Some nonparametric diagnostic statistical procedures and their asymptotic behavior

Gantner, M.

Publication date:
2010

Document Version
Publisher's PDF, also known as Version of record

Some Nonparametric Diagnostic Statistical Procedures and their Asymptotic Behavior

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit van Tilburg, op gezag van de rector magnificus, prof.dr. Ph. Eijlander, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op woensdag 8 september 2010 om 14.15 uur door

Maria Gantner

geboren op 24 juni 1982 te Dachau, Duitsland.

Promotor: prof.dr. John H.J. Einmahl

Commissieleden: prof.dr. Irène Gijbels
prof.dr. Laurens F.M. de Haan
prof.dr. Geurt Jongbloed
prof.dr. Jan R. Magnus
dr. Günther Sawitzki
prof.dr. Bas J.M. Werker

Thomas Stieltjes Institute
for Mathematics

The Uniform Song

There are continuous distributions,
discrete ones too.
Some are heavy tailed,
and some are skew.
There are logistics and chi squares,
but these we will scorn,
'Cause the loveliest of them all
is the Uniform.

G.R. Shorack and J.A. Wellner

First of all, I would like to express my sincere gratitude to John Einmahl for being such an outstanding supervisor to me. He is a wonderful teacher who can explain the most difficult concepts in a simple manner. John's kindness, enthusiasm, and the great respect with which he treats everyone around him made it such a joy to work with him. His knowledge and preciseness impressed me tremendously, his name and dedication opened many doors for me. I learned so much from him, scientifically and personally, and always left his office enlightened and happy. Thank you!

Furthermore I am very honored to have Irène Gijbels, Laurens de Haan, Geurt Jongbloed, Jan Magnus, Günther Sawitzki and Bas Werker as members of my PhD committee. I profited a lot from the cooperation with Günther on my first two papers, including a very nice stay in Heidelberg for a revision. In addition, he supported me, along with Regina Liu, on the job market. The last three years, Otilia Boldea often discussed my work with me and gave very valuable comments on my presentations. After finishing my thesis, Chen Zhou gave me the opportunity to work on another application of my first paper at the Erasmus University Rotterdam. Thank you all very much!

Tilburg University, CentER, and especially the department of Econometrics were an excellent working place. I could go to many nice conference and got

interesting courses to teach. I think part of the great atmosphere in our department was also due to our h.o.d. Bas Werker. Thank you!

Going to Tilburg was one of the best decisions in my life. Besides the aspect of work, this is also due to my social environment. I would like to take the opportunity to thank my colleagues and friends for the great time we had together and mention some key words: Antwerp, baking cookies, CA, cactus, cheese, Egyptian dinners, Freiburg, geen kapsalon voor mij, HvB, Istanbul, LG, München, my wonderful roommates, Nederlands voor medewerkers, New York, playing something like flamenco, Romania, running in the Warandebos, Spain, supervisor lunch, trophy wifes, Züri,...

Finally I thank my family: Mama und Papa, Anna und Stefan, and Marcel, for all their support and belief and much more.

Contents

Chapter 1

Introduction and Summary

In this thesis new nonparametric methods are introduced to explore uni- or bivariate data. These methods consist of two plots and one test:

1. A graphical method to depict probability distributions, the "shorth plot":
   a. Introduction to the method and applications (Chapter 2),
   b. Asymptotic behavior of the plot (Chapter 3).
2. A graphical method to depict regression data, the "Half-Half plot" (Chapter 4).
3. A test for spherical symmetry in an empirical likelihood framework (Chapter 5).

1.1 Introduction

In this section, we review the main statistical concepts of the thesis.

1.1.1 Nonparametric statistics

In Wolfowitz (1942) the term *nonparametric* was coined: "We shall refer to this situation [where a distribution is completely determined by the knowledge of its finite-parameter set] as the parametric case, and denote the opposite situation, where the functional forms of the distributions are unknown, as the nonparametric case". In contrast to parametric procedures, where the knowledge of the type of distribution is essential, nonparametric procedures have the great advantage that this knowledge is not required. As pointed out in Owen (2001), there is "indeed [...] no reason to suppose that a newly encountered set of data belongs to any of the well studied parametric families". In terms of statistical inference this means "to infer an unknown quantity while making as few assumptions as possible" (Wasserman, 2008). Misspecification in parametric procedures leads to inefficient estimates and the corresponding confidence intervals and tests can fail completely. Whereas when parametric assumptions hold perfectly true, nonparametric methods are in general only slightly less powerful than their parametric counterparts, see for example the textbook Kvam and Vidakovic (2007). Furthermore, according to Govindarajulu (2007), "nonparametric statistical procedures are widely used due to their simplicity, applicability under fairly general assumptions and robustness to outliers in the data".

The simulations and examples in this thesis are computed in `R`. A short introduction to especially nonparametric methods implemented in `R` can be found in Racine (2009); for a more general overview the reader is referred to the textbook Sawitzki (2009).

1.1.2 Graphical methods

Graphical methods are an important tool in statistics. As it is pointed out in Tufte (1983): "At their best, graphics are instruments for reasoning about quantitative information. Often the most effective way to describe, explore, and summarize a set of numbers - even a very large set - is to look at pictures

minimum	1^{st} quartile	median	3^{rd} quartile	maximum
11.29	15.67	18.24	21.65	34.55

Table 1.1: Summary of the quartiles of rent in Vienna's first district in 2010.

of those numbers". There exist many different graphical approaches for especially one- and two-dimensional data. A very broad overview is given in Tufte (1983), whereas du Toit et al. (1986) is more statistically oriented, and many graphical methods with R-code can be found in Theus and Urbanek (2009).

In the following, first some popular graphical methods for the case of one-dimensional data will be presented. The data are the ratios rent in euros per square meter (rent) in Vienna's first district (postal code 1010) in April 2010. The source of the data are websites of the most commonly used real estate agents in Vienna (www.alle-gemeinsam.at, www.derstandard.at, www.immoads.at, www.immobilien.net, www.immodirekt.at, www.oesterreich.kijiji.at, www.perconsult-immobilien.at, www.wohnnet.at) and the sample consists of 492 observations.

A first, exploratory quartile summary of the rent is given in Table 1.1. This information is also included in the boxplot, see Figure 1.1. Other commonly used methods to visualize one-dimensional data are the histogram (Figure

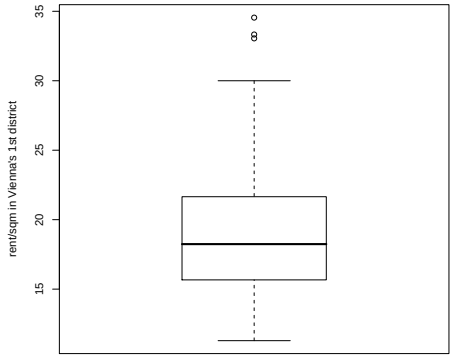

Figure 1.1: Boxplot of the rent in Vienna's first district in 2010.

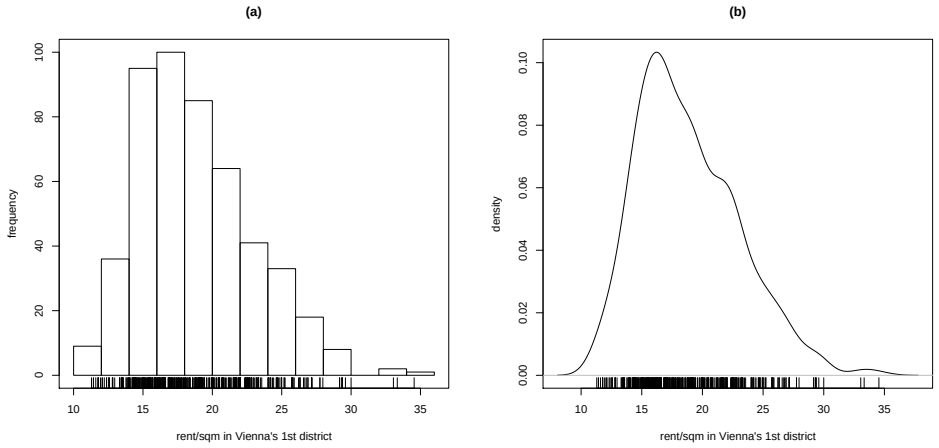

Figure 1.2: Histogram (a) and kernel density estimation (b) of the rent in Vienna's first district in 2010.

1.2(a)) and the kernel density estimate (Figure 1.2(b)). For both plots the choices of bins (histogram) and bandwidth (kernel density estimate) are crucial. Here the density estimation is plotted with a bandwidth of 1.08, which is calculated by Silverman's (1986) rule of thumb. The QQ-plot (Figure 1.3) is an exploratory tool to check a distributional assumption like, in this case, the normality of the data. In the QQ-plot, the order statistics of the data are plotted against the corresponding quantiles of a fitted normal distribution. Especially for smaller values of the rent, the plot drifts from the diagonal:

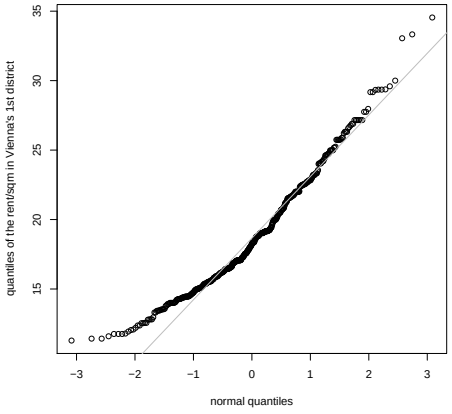

Figure 1.3: Normal QQ-plot of the rent in Vienna's first district in 2010.

The normality seems to be violated. With the shorth plot, we introduce a new approach to depict univariate mass concentration. A brief introduction to the shorth plot is given in Section 1.2.1.

For bivariate data, it is often important how the regressand Y can be explained in terms of a regressor X. In regression analysis, (X, Y) is then expressed in the standard form $Y = m(X) + \varepsilon$. To illustrate these graphical methods, we use the Nile data as reported in Cobb (1978), which are a popular

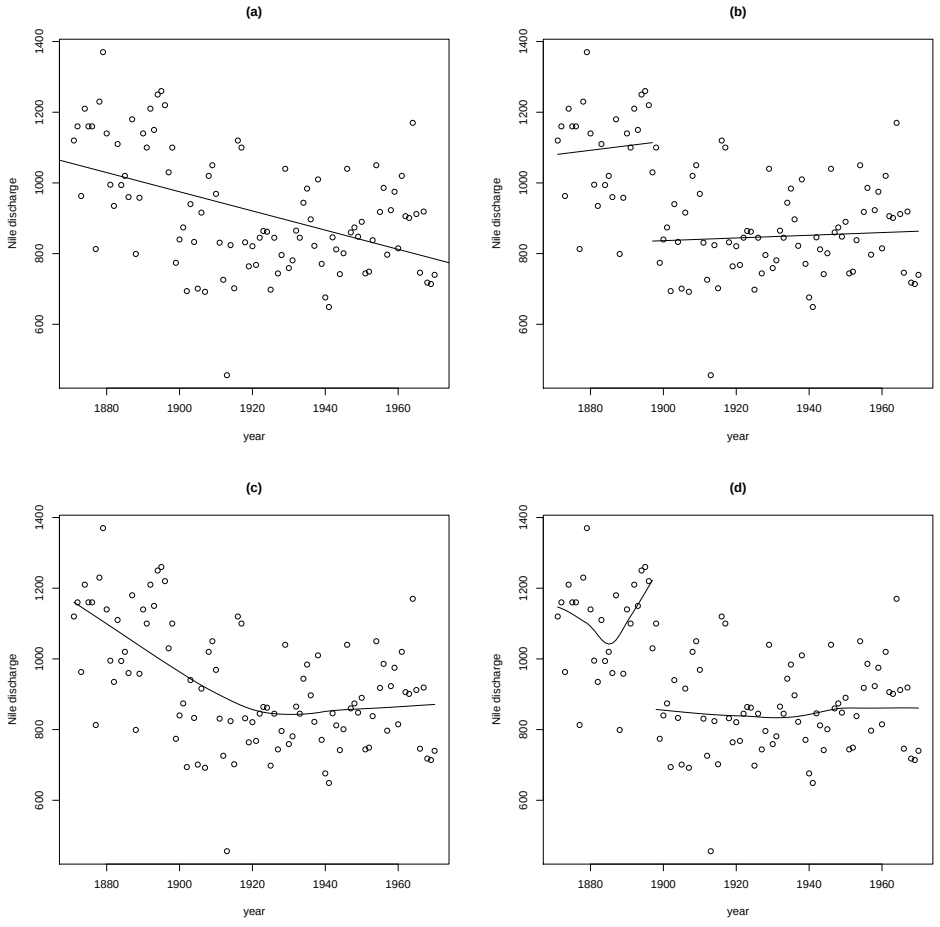

Figure 1.4: Scatter plot of the annual volume discharge of the Nile with estimated regression curves. Linear regression in (a) and (b); in panel (b) the break in 1897 is taken into account. In panels (c) and (d), the curve is fitted with locally-weighted polynomial regression; in panel (d) the break is taken into account.

dataset used in the context of change-point estimation. Figure 1.4 shows the annual volume of discharge ($10^{10}m^3$) from the Nile river for each year from 1871 through 1970. The regressor is in this example deterministic (the years 1871-1970). In Cobb (1978) it is assumed that the ε's are independent. There a change-point in 1897 is found utilizing parametric methods. Meteorological studies confirm this change. According to Kraus (1955), the rainfall decreased in most regions abruptly at the end of the 19th century, which was due to a narrowing of the rainfall belt and a shortening of the wet seasons.

The most standard form of regression is linear regression, see Figure 1.4 (a). In Figure 1.4 (b) two linear regression lines, left and right of the break in 1897, are fitted. Locally-weighted polynomial regression, as introduced in Cleveland (1979), allows for a more flexible fit, see Figures 1.4 (c,d), where the curves are fitted left and right of the break in 1897. With the Half-Half plot we shall introduce in this thesis a new graphical method for detecting the main features of regression curves. A short introduction to the Half-Half plot can be found in Section 1.2.2.

1.1.3 Hypotheses tests

Nonparametric tests are often goodness-of-fit tests. They assess how well a statistical model fits a set of observations. Well-known goodness-of-fit tests are the Kolmogorov-Smirnov test and the Anderson-Darling test. Let $X_1, \ldots, X_n$, $n \in \mathbb{N}$, be i.i.d. random variables with common (unknown) continuous distribution function (df) F and empirical df F_n. The null hypothesis is whether this df is of the exactly specified form F_0. The hypotheses for a two-sided test are

$$\begin{aligned} H_0 &: F(x) = F_0(x) \quad \text{for all } x \in \mathbb{R}, \\ H_1 &: F(x) \neq F_0(x) \quad \text{for at least one } x \in \mathbb{R}. \end{aligned}$$

The Kolmogorov-Smirnov statistic is defined as

$$K_n = \sup_{x \in \mathbb{R}} |F_n(x) - F_0(x)|,$$

and is distribution-free. Under H_0, $K_n \xrightarrow{\mathbb{P}} 0$; furthermore,

$$\sqrt{n}K_n \xrightarrow{d} \sup_{0<t<1} |B(t)|,$$

where B denotes a standard Brownian bridge. The null hypothesis is rejected for large values of $\sqrt{n}K_n$, see for example Shorack and Wellner (1986), p. 143. A Kolmogorov-Smirnov test may also be used in the two-sample case to test whether the two underlying one-dimensional df's differ.

The Anderson-Darling (1952) statistic is defined as

$$W_n^2 = n \int_{-\infty}^{\infty} [F_n(x) - F_0(x)]^2 \, \psi [F_0(x)] \, dF_0,$$

where $\psi \geq 0$ denotes a preassigned weight function. The null hypothesis is again rejected for large values of W_n^2, see Crawford et al. (1990). The asymptotic distribution of W_n^2 for $\psi(t) = \frac{1}{t(1-t)}$ is derived in Anderson and Darling (1952). For $\psi \equiv 1$, W_n^2 is also called the Cramèr-von-Mises test statistic.

Instead of testing for a fixed F_0, we can also test for a parametric model. Then we compare F_n with the parametric df where the parameters are estimated from the data. In that case, the critical values of both the Kolmogorov-Smirnov and the Anderson-Darling statistic, as introduced above, are invalid. For some special distributions, such as the normal or exponential distribution, these critical values are given in Shorack and Wellner (1986), p. 239.

Furthermore, instead of testing for a parametric model, there also exist tests for a larger, nonparametric null hypothesis. They check for a general feature, such as symmetry, see for example the tests introduced in Einmahl and McKeague (2003). In this thesis we shall develop a test for bivariate spherical symmetry, briefly presented in Section 1.2.3.

1.1.4 Asymptotic statistics

"Why asymptotic statistics? The use of asymptotic approximations is twofold. First, they enable us to find approximate tests and confidence regions. Second, approximations can be used theoretically to study the quality (efficiency) of statistical procedures" (van der Vaart, 1998).

The exact distributions of statistics are often unknown, whereas asymptotic approximations are available. Let $X_1, \ldots, X_n$, $n \in \mathbb{N}$, be i.i.d. random variables with sample mean $\bar{X}_n$ and sample standard deviation S_n. The t-test is a test on the mean $\mathbb{E}(X) = \mu$, we test $H_0 : \mu = \mu_0$ for a given μ_0. In the case that the data come from a normal distribution, the distribution of the t-statistic $T_n = \sqrt{n}\left(\bar{X}_n - \mu_0\right)/S_n$ is known under H_0: It is t-distributed with $n-1$ degrees of freedom. In general, if the underlying distribution is not normal, but has finite variance, the following approximation by the standard normal df Φ holds:

$$\sup_{x\in\mathbb{R}} \left| P\left(\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{S_n} \leq x\right) - \Phi(x) \right| \to 0 \qquad \text{as } n \to \infty, \text{ for all } \mu \in \mathbb{R}. \quad (1.1)$$

When $\mu = \mu_0$, the df of T_n can therefore be approximated by Φ and H_0 gets rejected at level α if $|T_n| \geq \Phi^{-1}(1-\alpha/2)$. From (1.1) we also immediately obtain an asymptotic confidence interval for μ: $\bar{X}_n \pm \Phi^{-1}(\alpha/2)S_n/\sqrt{n}$.

In this thesis, we introduce three nonparametric statistical methods (cf. Section 1.2) and discuss their asymptotic behavior. For the test in Chapter 5 we present the limiting null distribution of the test statistic and hence we can find asymptotic critical values. In Chapter 4 we employ the limiting distribution to calculate a band which gives a standard to assess the relative magnitude of the HH-statistic. Asymptotic statistics also provides the rate of convergence. The typical speed of convergence of kernel density estimators is $n^{-2/5}$. In Chapter 3 we show that the shorth plot converges at rate $n^{-1/2}$.

1.2 Summary

In this section, the three main parts of this thesis are summarized.

1.2.1 The shorth plot

The "shorth" is originally the *short*est interval covering *h*alf of the distribution. For the shorth plot, a generalized version is used, namely the *length* of the shortest interval containing fraction α of the distribution and a given point $x \in \mathbb{R}$. This length is plotted against x for a selection of α's. The shorth plot visualizes probability mass concentration. The classical statistical approach for this task focuses on density estimators and their visual representations, such as kernel density plots. In kernel density estimation, the bandwidth choice is crucial. The shorth plot, however, discloses the features of the distribution without bandwidth selection. For the exact definition of the shorth plot see Chapter 2.

Figure 1.5 depicts an example of the shorth plot. The data are a sample of the ratios rent/m^2 in Vienna's first district in April 2010, as introduced in Section 1.1.2. The different coverage levels α reveal different information

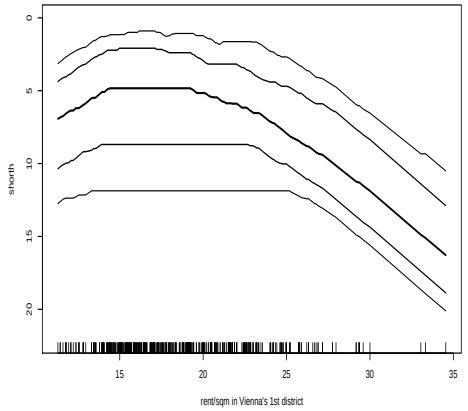

Figure 1.5: Shorth plot of the rent in Vienna's first district in 2010, with coverage levels (from top to bottom) $\alpha = 0.125, 0.25, 0.5, 0.75$, and 0.875.

about the concentration of mass. Small levels give information about the local behavior; they particularly indicate modes. Higher levels show the skewness of the overall distribution shape. It is advised to draw several coverage levels in one picture, which then reveal the different characteristics of the data. The shorth plot in Figure 1.5 immediately reveals that the data are right-skewed. From $\alpha = 0.125$, a clear mode around 16.5 can be detected, and a less distinct mode around 19 becomes apparent. A third mode around 23 has clearly less height. The second and third mode can also become visible in kernel density estimation with a smaller choice of bandwidth than in Figure 1.2(b).

For the asymptotic behavior of the shorth plot, we first regard the "shorth plot process" which is a quantile-type process. We can show that this process converges in distribution under natural conditions uniformly in $\alpha \in [\eta, 1-\eta]$, $\eta \in (0, \frac{1}{2})$, and $x \in \mathbb{R}$ with rate $1/\sqrt{n}$ to a limiting process, see Chapter 3. For the uniform convergence in $\alpha \in (0, 1)$ and $x \in \mathbb{R}$, we need two more assumptions. The proof includes a lemma that generalizes the well-known Vervaat (1972) lemma to a collection of functions.

1.2.2 The Half-Half plot

The Half-Half (HH) plot is a new, nonparametric, computationally fast method for detecting the main features of regression curves. The empirical HH-plot

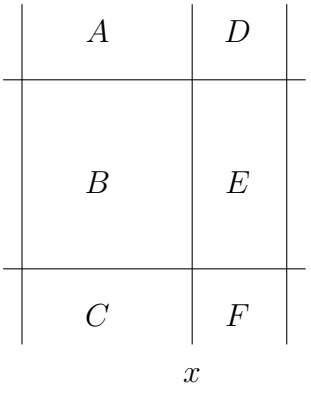

Figure 1.6: Vertical strip localized at x used for the construction of the HH-plot.

counts observations in the lower and upper quarter of a vertical strip that moves horizontally over the scatter plot. In Figure 1.6 a strip is shown, localized at x. It contains $\lceil \alpha n \rceil$ observations in $A \cup B \cup C$ as well as in $D \cup E \cup F$, with $\lceil \cdot \rceil$ denoting the ceiling function. This strip is divided up horizontally such that $B \cup E$ contains the central 50% of the observations in the strip. Therefore $A \cup D$ and $C \cup F$ each contain 25% of the $2\lceil \alpha n \rceil$ observations. For the HH-plot, the (standardized) number of observations in $C \cup D$ is plotted against the localization x. The HH-plot is, similar to the shorth plot, a tool for exploratory diagnostics. It detects especially jumps and trends of the regression curve; the jumps would typically be forced into a smoothed picture by the common methods, see Figure 1.4 (a,c). For the exact definition of the HH-plot consult Chapter 4.1.

For a first picture of the HH-plot, we use again the Nile data, as discussed in Section 1.1.2. This dataset was first used in Carlstein (1988) to illustrate *nonparametric* change-point estimation. In previous publications utilizing this dataset, a change-point in 1897 was found. Also the HH-plot with $\alpha = 0.2$, Figure 1.7, displays this change clearly, indicated by an exceedance of a horizontal band. This band gives a standard to assess the relative magnitude of the HH-plot. Because the HH-plot is below the band in 1895–1897, we can conclude that there is an abrupt change downwards of the regression curve

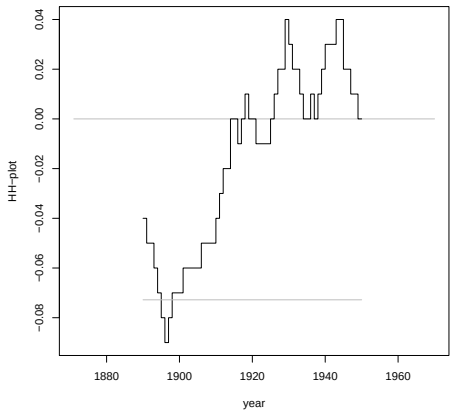

Figure 1.7: HH-plot of the annual volume discharge of the Nile, with $\alpha = 0.2$.

which can be located at 1896. In the second half of the period under consideration, positive values of the HH-plot indicate that the regression curve is slightly increasing.

1.2.3 Testing for spherical symmetry using localized empirical likelihood

In Chapter 5, a test for spherical symmetry in $\mathbb{R}^2$ is developed within an empirical likelihood framework. Empirical likelihood is a nonparametric counterpart of the parametric likelihood ratio and enjoys similar features. In contrast to parametric likelihood methods, the data are used directly, without the assumption of a certain underlying family of distributions. We employ the localized empirical likelihood approach, see Einmahl and McKeague (2003). The main idea is that a functional equation is 'split up' in infinitely many pointwise equations and that standard empirical likelihood is used to deal with these pointwise constraints, after which the infinitely many likelihood ratios are considered simultaneously as a stochastic process.

In this introduction we consider the simple null hypothesis $H_0 : F = F_0$, where F_0 is a given, completely specified, continuous distribution function. Given are i.i.d. random variables $X_1, \ldots, X_n$, $n \in \mathbb{N}$ that have df F. Next we localize the null hypothesis in x, i.e. we consider $H_{0,x} : F(x) = F_0(x)$. Define the localized empirical likelihood as

$$R(x) \;=\; \frac{\sup\{L(\tilde{F}) : \tilde{F}(x) = F_0(x)\}}{\sup\{L(\tilde{F})\}},$$

where $L(\tilde{F}) = \prod_{i=1}^{n} \Big(\tilde{F}(X_i) - \tilde{F}(X_i-)\Big)$. The supremum in the denominator is attained by taking $\tilde{F}$ equal to the empirical distribution function F_n of the X_i's. In the numerator, the maximum is obtained when $\tilde{F}$ puts mass $\frac{F_0(x)}{nF_n(x)}$ on each observation up to x and mass $\frac{1-F_0(x)}{n(1-F_n(x))}$ on each observation above x.

This leads to

$$R(x) \;=\; \frac{\left(\frac{F_0(x)}{nF_n(x)}\right)^{nF_n(x)} \left(\frac{1-F_0(x)}{n(1-F_n(x))}\right)^{n((1-F_n(x))}}{\left(\frac{1}{n}\right)^n},$$

and hence

$$-2\log R(x) \;=\; -2nF_n(x)\log\frac{F_0(x)}{F_n(x)} - 2n(1-F_n(x))\log\frac{1-F_0(x)}{1-F_n(x)}.$$

Applying a Taylor expansion, given $0 < F_0(x) < 1$, we get

$$-2\log R(x) \;=\; \frac{n\left(F_n(x)-F_0(x)\right)^2}{F_0(x)\left(1-F_0(x)\right)} + o_P(1) \xrightarrow{d} \chi^2_1 \tag{1.2}$$

under $H_{0,x}$. This is a special case of a nonparametric version of Wilks' theorem, cf. Owen (2001). Considering the infinitely many localized likelihoods simultaneously, we take the integral of (1.2) with respect to F_0 and obtain the test statistic

$$S_n \;=\; -2\int_{-\infty}^{\infty} \log R(x) dF_0(x).$$

The test statistic is distribution-free and the critical values can be approximated by simulation. It can be shown, using empirical process theory, that

$$S_n \xrightarrow{d} \int_0^1 \frac{B^2(t)}{t(1-t)}dt$$

under H_0, where B is a standard Brownian bridge.

To test for bivariate spherical symmetry around the origin, we use its stochastic representation to construct the test. The random variable X is spherically symmetric if and only if

$$X \overset{d}{=} SZ,$$

for some random variable $S \geq 0$ independent of the random vector Z and Z is uniformly distributed on the unit circle. With the approach introduced above, a test statistic and its limiting null distribution are derived in Chapter 5.

Chapter 2

The shorth plot

[Based on joint work with J.H.J. Einmahl and G. Sawitzki, *The shorth plot*, Journal of Computational and Graphical Statistics (2010), 19, 62–73.]

Abstract: The shorth plot is a tool to investigate probability mass concentration. It is a graphical representation of the length of the shorth, the shortest interval covering a certain fraction of the distribution, localized by forcing the intervals considered to contain a given point x. It is easy to compute, avoids bandwidth selection problems and allows scanning for local as well as for global features of the probability distribution. The good performance of the shorth plot is demonstrated through simulations and real data examples. These data as well as an `R`-package for computation of the shorth plot are available online.

Keywords: Data analysis, distribution diagnostics, probability mass concentration.

2.1 Introduction

Using exploratory diagnostics is one of the first steps in data analysis. Here graphical displays are essential tools. For detecting specific features, specialized displays may be available. For example, if there is a model distribution F which needs to be compared with, from a mathematical point of view the

empirical distribution F_n is a key instrument, and its graphical representations by means of PP-plots

$$x \mapsto (F(x), F_n(x))$$

or QQ-plots

$$\alpha \mapsto \left(F^{-1}(\alpha), F_n^{-1}(\alpha)\right)$$

are tools of first choice. If we consider the overall scale and location, box & whisker plots are a valuable tool. The limitation of box & whisker plots is that they give a global view which ignores any local structure. In particular, they are not an appropriate tool for analyzing the modality of a distribution. In this case, more specialized tools are needed, such as for example the silhouette and the excess density plot, both being introduced in Müller and Sawitzki (1991).

While we have some instruments for specific tasks, the situation is not satisfactory if it comes to general purpose tools. PP-plots and QQ-plots need considerable training to be used as diagnostic tools, as they do not highlight the qualitative features of the data.

The classical statistical approach is to focus on the density in contrast to the distribution function, which leads to density estimators and their visual representations, such as histograms and kernel density plots. These, however, introduce another complexity, such as the choice of cut points or bandwidth choice. The qualitative features revealed or suggested by density estimators may critically depend on bandwidth choice. A large bandwidth tends to lead to oversmoothing and hides features and a small bandwidth is prone to undersmoothing and will produce sample artefacts. In Silverman (1981) these properties are exploited to develop a test for multimodality. The need to consider various bandwidths and cutpoints for histograms or kernel density estimators is widely recognized, but the practical application is limited, since the plots for different bandwidths overlap and do not give a possibility for a multiscale representation.

The SiZer approach in Chaudhuri and Marron (1999) tries a way out by using nominal test levels for varying bandwidths. There is considerable literature

on good or optimal choice of bandwidth. However, most of this literature concentrates on kernel methods or histograms as density estimators and typically uses optimality criteria based on squared errors. In applications, however, these methods are more often used to inspect data for distribution features, and unfortunately a criterium like the mean integrated squared error does not translate easily to any feature of interest.

The shorth plot, however, discloses the features of the distribution without bandwidth selection due to its monotonicity property (see the end of this section). In contrast to density estimation, it is advised to plot several coverage levels in one picture, which then reveal the different characteristics of the data. Moreover, estimating a density is a more specific task than understanding the shape of a density. Density estimation based methods are prone to pay for bandwidth selection in terms of slow convergence or large fluctuation, or disputable choices of smoothing. Hence the shorth plot, which displays the qualitative features of the distribution without estimating the density, can have a faster rate of convergence than density estimators.

We will use the length of the shorth to analyze the qualitative shape of a distribution. Originally, the shorth is the shortest interval containing half of the distribution; more generally, the α-shorth is the the shortest interval containing fraction α of the distribution. The shorth was introduced in the Princeton robustness study as a candidate for a robust location estimator, using the mean of a shorth as an estimator for a mode, see Andrews et al. (1972). As a location estimator, it performs poorly; it has an asymptotic rate of only $n^{-1/3}$, with non-trivial limiting distribution, as shown in Andrews at al. (1972), p. 50, or Shorack and Wellner (1986), p. 767. Moreover, the shorth interval is not well defined, since there may be several competing intervals. The length of the shorth however is a functional which is easy to estimate and it gives a graphical representation which is easy to interpret. As pointed out in Grübel (1988), the length of the shorth has a convergence rate of $n^{-1/2}$. In this paper, we extend the definition of the length of the shorth to supply localization. We will vary the coverage α, and hence allow for multi-scale analysis. Thus the global estimator is extended into a tool for local and global

diagnostics. The shorth plot is already valuable for sample sizes of $n = 50$ (for coverage levels above $\alpha = 0.1$).

The paper is organized as follows. In the remainder of this section we present the definition and elementary properties of the localized length of the shorth. In Section 2.2 we define the shorth plot, the central object of this paper, and show its use. In Section 2.3 we study several real data examples. The paper is completed by a discussion section.

In order to be more explicit we specify our setup and notation. Let $X_1, \ldots, X_n$, $n \geq 1$, be independent random variables with common distribution function F. Let P be the probability measure corresponding to F. Let $\mathbb{I} = \{[a, b] : -\infty < a < b < \infty\}$ be the class of closed intervals and let $\mathbb{I}_x = \{[a, b] : -\infty < a < b < \infty,\ x \in [a, b]\}$ be the class of closed intervals that contain $x \in \mathbb{R}$. Define the empirical measure P_n on the Borel sets $\mathbb{B}$ on $\mathbb{R}$ by

$$P_n(B) = \frac{1}{n} \sum_{i=1}^{n} 1_B(X_i), \quad B \in \mathbb{B},$$

where 1_B denotes the indicator function. Let $|\cdot|$ denote Lebesgue measure.

Definition 2.1. The length of the shorth at point $x \in \mathbb{R}$ for coverage level $\alpha \in (0, 1)$ is

$$S_\alpha(x) = \inf\{|I| : P(I) \geq \alpha, I \in \mathbb{I}_x\}.$$

By taking $\inf_x S_{0.5}(x)$, we get the length of the shorth as originally defined. The definition in terms of a theoretical probability P has an immediate empirical counterpart, the empirical length of the shorth

$$S_{n,\alpha}(x) = \inf\{|I| : P_n(I) \geq \alpha, I \in \mathbb{I}_x\}.$$

To get a picture (see Figure 2.1) of the optimization problem behind the length of the shorth, we consider the bivariate function

$$(a, b) \longmapsto \big(\, |I| \,, P\,(I)\,\big) \qquad \text{with } I = [a, b], \text{ where } a < b.$$

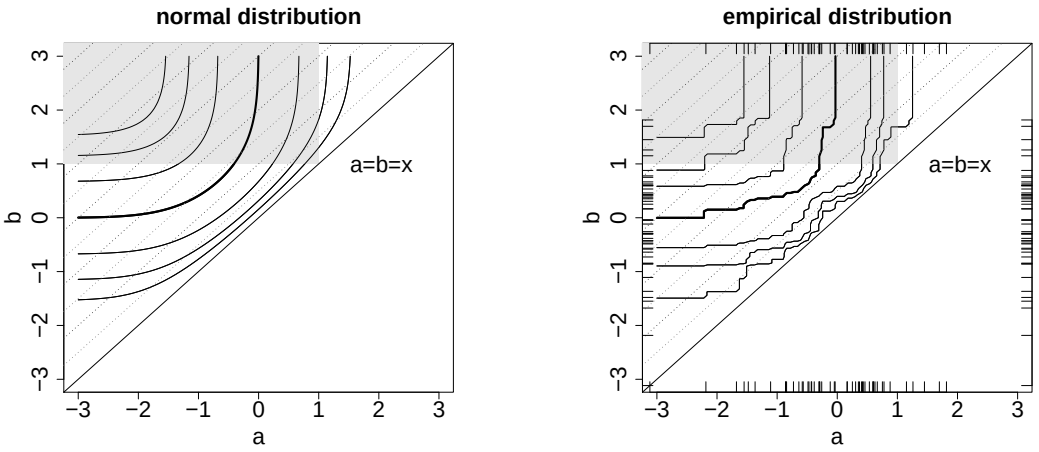

Figure 2.1: The length of the shorth as an optimization problem: minimize $|[a,b]|$ under the restriction $P([a,b]) \geq \alpha$. Localizing at x restricts the optimization to the quadrant top left of (x,x). The solid lines are the level curves for $P([a,b]) = const = \alpha$, for $\alpha = 0.9375,\ 0.875,\ 0.75,\ 0.5,\ 0.25,\ 0.125,\ 0.0625$ from top to bottom. The bold line depicts $\alpha = 0.5$. The dashed lines are the level curves for $|[a,b]| = const$. For the empirical distribution, a sample of size 100 from a standard normal distribution is taken.

This is defined on the half space $\{(a,b) : a < b\}$ above the diagonal. The level curves of $|I|$ are parallel to the diagonal. The level curves of $P(I)$ depend on the distribution. The α-shorth minimizes $|I|$ in the area above the level curve at level α, i.e. $P(I) \geq \alpha$. In the empirical version, the level curves of $P(I)$ are replaced by those of of $P_n(I)$. In Figure 2.1, the theoretical curves for the Gaussian distribution and for a Gaussian sample are shown. Localizing the α-shorth at a point x restricts optimization to the (grey) top left quadrant anchored at (x,x).

Let the distribution function F be absolutely continuous with density f. Assume there exist $-\infty \leq x_* < x^* \leq \infty$ such that $f(x) > 0$ on $\mathcal{S} = (x_*, x^*)$ and $f(x) = 0$ outside $\mathcal{S}$; also assume that f is uniformly continuous on $\mathcal{S}$. As a consequence, F is strictly increasing on $\mathcal{S}$ and f is bounded. We have the following elementary properties concerning $S_\alpha(x)$.

• *Minimizing intervals:* For every α and x, there exists an interval I with length $S_\alpha(x)$ such that $x \in I$ and $P(I) = \alpha$.

- *Continuity:* For all α, $|S_\alpha(x) - S_\alpha(y)| \leq |x - y|$. Moreover, the function

$$(x, \alpha) \mapsto S_\alpha(x)$$

is continuous as a function of two variables.

- *Monotonicity:* For all x,

$$\alpha \mapsto S_\alpha(x)$$

is strictly increasing in α.

- *Invariance:* For all α,

$$x \mapsto S_\alpha(x)$$

is invariant under shift transformations and equivariant under scale transformations, that is when we apply a transformation $u' = cu + d$ (for some constants $c > 0, d$), then the new $S'_\alpha(x')$ satisfies

$$S'_\alpha(x') = cS_\alpha(x),$$

with $x' = cx + d$.

Denote the j-th order statistic by $X_{(j)}$; $X_{(0)} = -\infty, X_{(n+1)} = \infty$. For computing the empirical length of the shorth, observe that $S_{n,\alpha}(x)$ can be interpolated from $S_{n,\alpha}(X_{(j)})$ and $S_{n,\alpha}(X_{(j+1)})$ where j is such that $X_{(j)} \leq x < X_{(j+1)}$. Therefore we can focus on computing $S_{n,\alpha}(X_i)$. Write $k_\alpha = \lceil n\alpha \rceil - 1$, with $\lceil \cdot \rceil$ the ceiling function. Then we simply have

$$S_{n,\alpha}(X_i) = \min\{X_{(j+k_\alpha)} - X_{(j)} : \ 1 \leq j \leq i \leq j + k_\alpha \leq n\}.$$

We can further reduce the complexity of our problem because $S_{n,\alpha}(X_i)$ can be easily related to $S_{n,\alpha}(X_{i-1})$ through a stepwise algorithm. The resulting algorithm has a linear complexity in n.

2.2 The Shorth Plot

Definition 2.2 (Sawitzki, 1994)**.** The shorth plot is the graph of the functions

$$x \mapsto S_\alpha(x),\ x \in \mathbb{R}$$

for (all or) a selection of coverages α.

The empirical shorth plot is the graph of

$$x \mapsto S_{n,\alpha}(x),\ x \in \mathbb{R}.$$

Mass concentration can be represented by the graphs of $x \mapsto S_\alpha(x)$ and $x \mapsto S_{n,\alpha}(x)$, see Figure 2.2. A *small* length of the shorth signals a large probability mass. In contrast, *large* values of the density indicate a large probability mass. Therefore, to make the interpretation of the shorth plot easier, we will in the sequel use a downward orientation of the vertical axis so that it is aligned with the density plot.

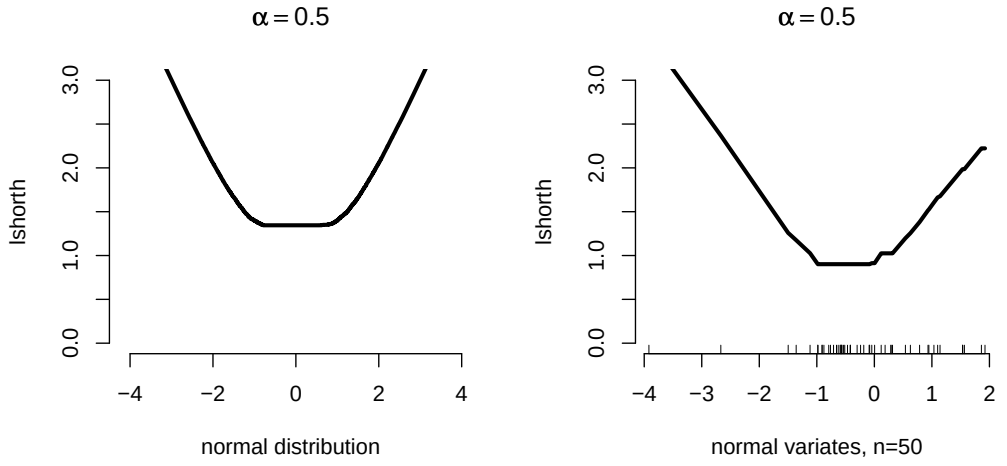

Figure 2.2: Short plot and empirical shorth plot for a sample of 50 standard normal random variables for $\alpha = 0.5$. Note that different scales are used.

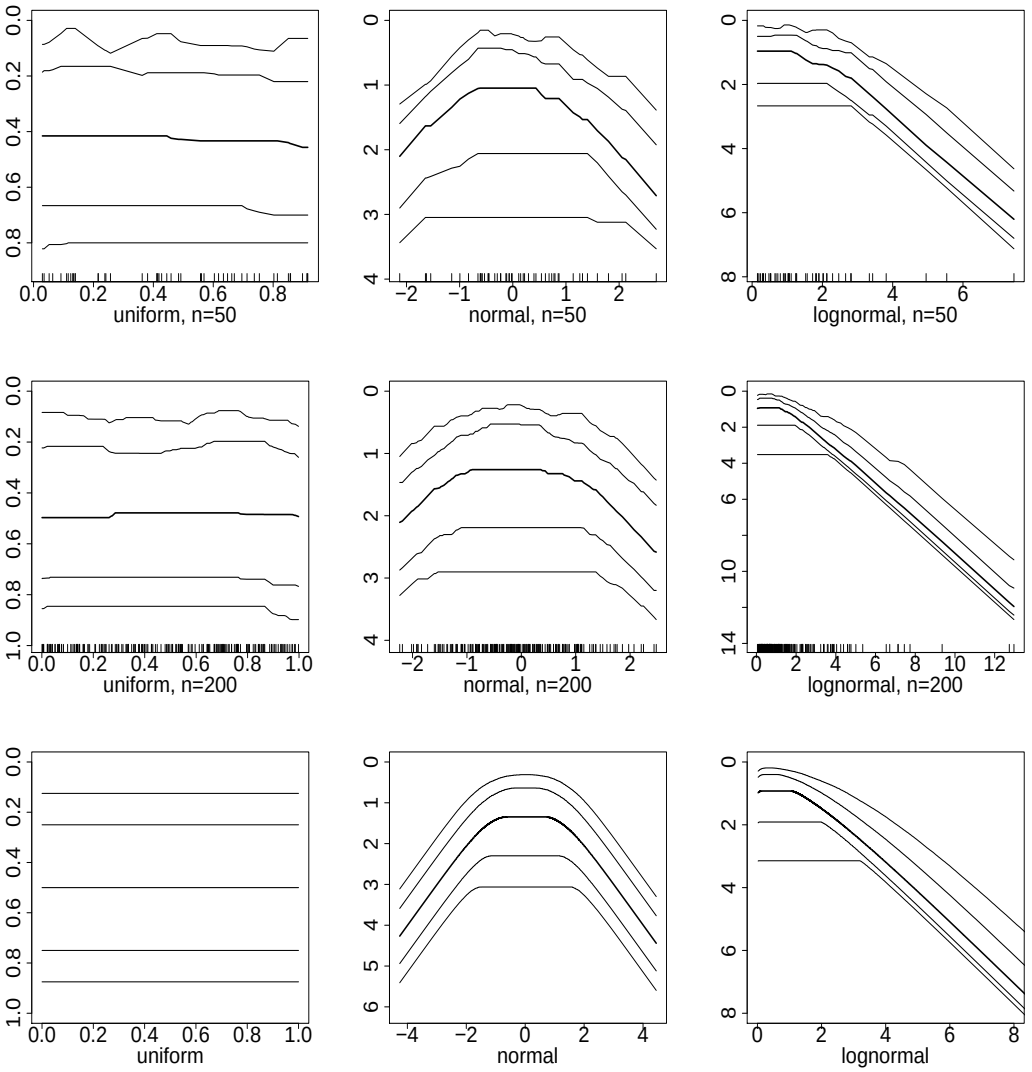

Figure 2.3: Shorth plots for a uniform, a normal, and a log-normal distribution for sample sizes 50 and 200 and the theoretical ones, for coverage levels (from top to bottom) $\alpha = 0.125, 0.25, 0.5, 0.75, 0.875$. Note that the vertical axes have a downward orientation and that different scales are used.

Figure 2.3 shows the empirical shorth plots for a uniform, a normal and a log-normal distribution for sample sizes 50 and 200 and the theoretical ones. Varying the coverage level α gives a good impression of the mass concentration. Small levels give information about the local behavior, in particular near modes. Higher levels give information about skewness of the overall distribution shape. The high coverage levels show the range of the distribution. A "dyadic" scale for α, e.g., 0.125, 0.25, 0.5, 0.75, 0.875 is a recommended choice. The Monotonicity property (Section 2.1) allows the multiple scales to be displayed simultaneously without overlaps, thus giving a multi-resolution image of the distribution.

Let ϕ denote the standard normal density. Then

$$f(x) = \begin{cases} 0.05 & \text{for } |x| \le 0.13 \\ 0.987\,[\phi(x) + \phi(|x| - 0.13)] & \text{for } 0.13 < |x| < 0.26 \\ 0.987\,\phi(x) & \text{for } |x| \ge 0.26 \end{cases} \tag{2.1}$$

is an example of a density where the shorth plot outperforms other common methods. The plot of this density is depicted in Figure 2.4. We consider a random sample of size 500. The fixed bandwidth kernel density estimate with an "optimal" bandwidth (calculated with Silverman's rule of thumb, Silverman (1986, p.48, equation (3.31)) suggests a density close to a standard

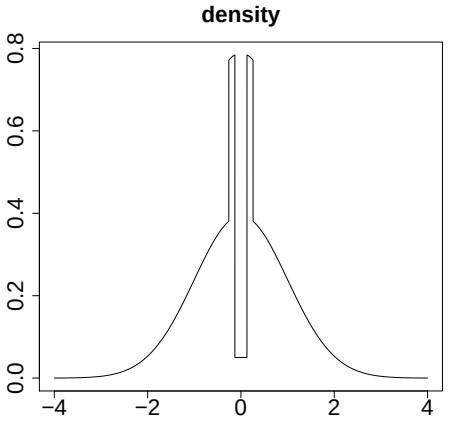

Figure 2.4: Plot of the density f in (2.1).

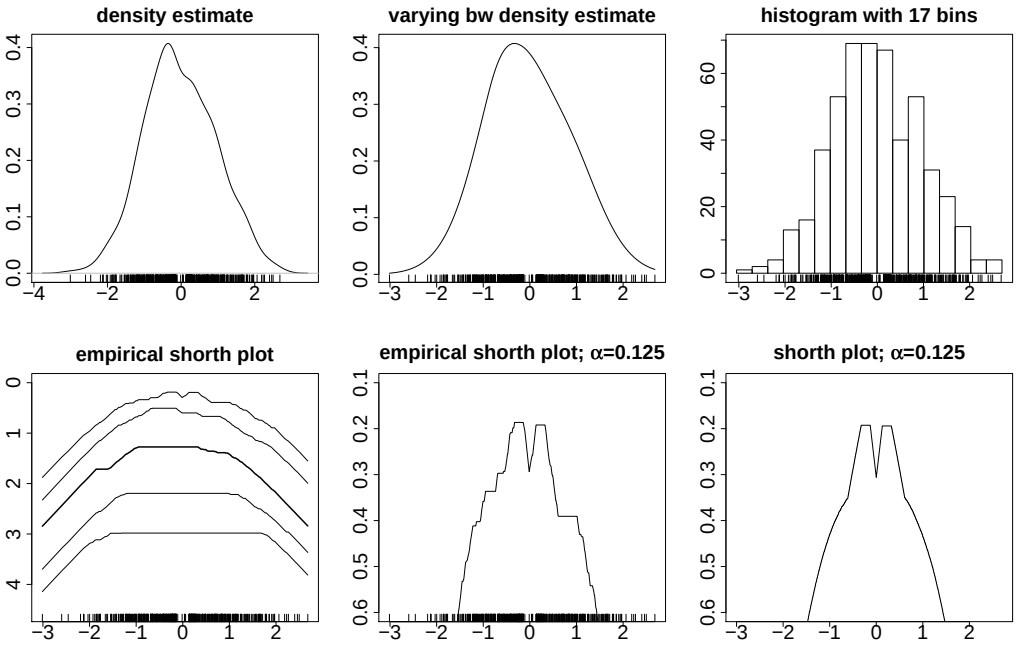

Figure 2.5: Density estimators (upper panels) and shorth plot (with estimators; lower panels) of f in (2.1); sample size $n = 500$. *Upper panels:* Kernel density estimate of f with "optimal" bandwidth $h = 0.2524$; kernel density estimate of f with varying bandwidth; histogram with 17 bins. *Lower panels:* Empirical shorth plot of the sample with coverage levels (from top to bottom) $\alpha = 0.125,\ 0.25,\ 0.5,\ 0.75,\ 0.875$; empirical shorth plot of the sample for $\alpha = 0.125$ (note the fine scale); shorth plot of f for $\alpha = 0.125$.

normal, see Figure 2.5 (upper-left panel). Estimating the density with a varying bandwidth results in a plot even more similar to the standard normal, see the upper-middle panel of Figure 2.5. (This kernel density estimator is calculated according to Loader (1999, Chapters 3 and 5), using the default nearest neighbor fraction.) Also a histogram with 17 bins (upper-right panel of Figure 2.5) sheds no light on the true shape of the density.

The shorth plot, however, detects as well the spikes of the density at $0.13 < |x| < 0.26$ as the dip at $|x| \leq 0.13$, and clearly outperforms both density estimators. This behavior is revealed by the small coverage levels; in order to make it more visible, we plot in the lower-middle panel of Figure 2.5 only the empirical shorth plot for coverage level $\alpha = 0.125$ on a finer scale. For

comparison, the theoretical shorth plot for $\alpha = 0.125$ is given in the lower-right panel of Figure 2.5. Increasing coverage levels in the shorth plot show the symmetry and the range of the distribution, respectively.

2.3 Examples

In this section, we use the shorth plot to depict several real data examples and explain how various features of the distribution can be "read" from this plot.

2.3.1 Annual maximum river discharges of the Meuse river

The annual maximum discharges of the Meuse river from 1911 through 1995 at Borgharen in The Netherlands are an example of right skewed unimodal data. The data are extensively studied in Beirlant et al. (2004), using extreme value theory. The shorth plot shows clearly the unimodality and the right skewness.

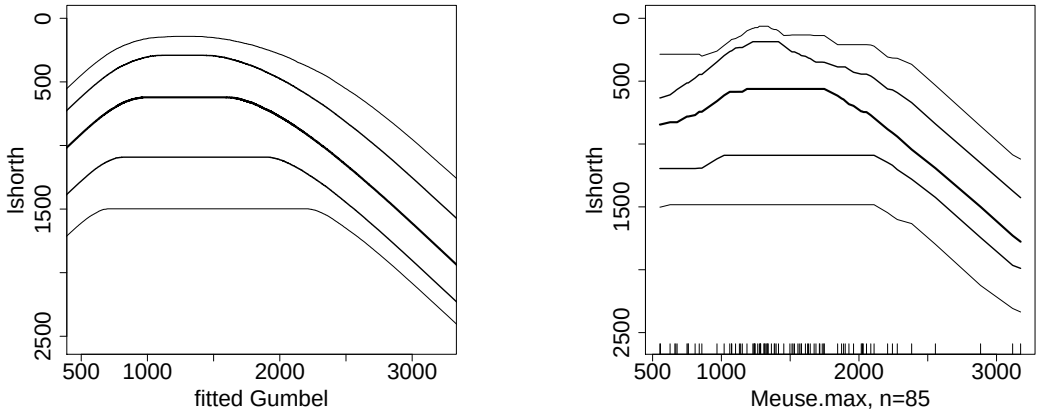

Figure 2.6: Shorth plots with coverage levels (from top to bottom) $\alpha = 0.125$, 0.25, 0.5, 0.75 and 0.875: (left) Fitted Gumbel distribution. (right) Annual maximum discharges of the Meuse river from 1911 through 1995.

The mode (about 1300) is indicated by small values of the shorth for the small coverage levels, see Figure 2.6 (right). In Beirlant et al. (2004) it is shown that the Gumbel distribution fits the data quite well and hence supports its common use in hydrology for annual river discharge maxima. For comparison, we picture the shorth plot of the fitted Gumbel distribution in the left panel of Figure 2.6. The right skewness is in both plots clearly visualized.

2.3.2 Old Faithful Geyser

As a second example, we use the eruption durations of the Old Faithful geyser. The data are just one component of a bivariate time series data set. Looking at a one dimensional marginal distribution ignores the process structure. However, these data have been used repeatedly to illustrate smoothing algorithms like (fixed bandwidth) kernel density estimators (Figure 2.7, left), and we reuse it to illustrate our approach (Figure 2.7, right). This is a good natured data set showing two distinct nodes with sizeable observation counts, and some overall skewness. The high coverage levels of the shorth plot ($\alpha = 0.75,\ 0.875$) just show the overall range of the data. The 0.5 level indicates a pronounced

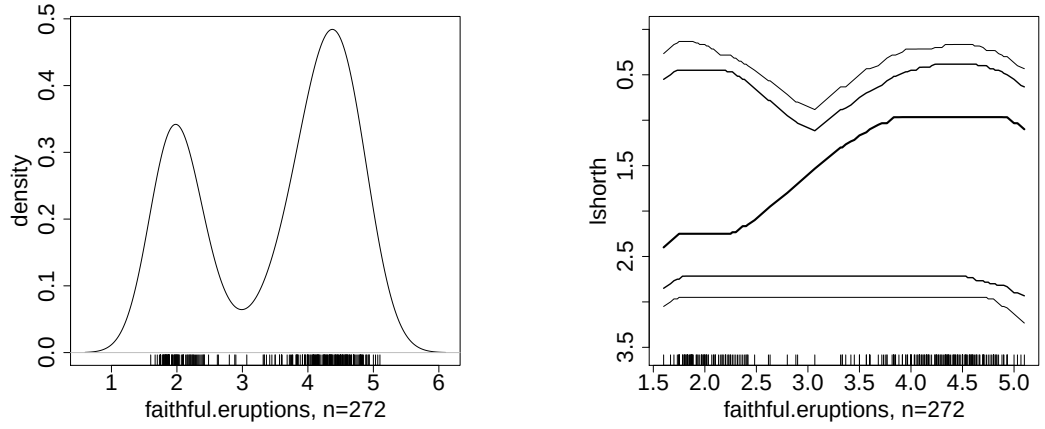

Figure 2.7: Eruption durations of the Old Faithful geyser: (left) Kernel density estimate ("optimal" bandwidth $h = 0.3348$). (right) Shorth plot with coverage levels (from top to bottom) $\alpha = 0.125,\ 0.25,\ 0.5,\ 0.75$ and 0.875.

skewness, whereas the small levels reveal two modes of approximately the same height (which the density estimate fails to show). The multi-scale property of the shorth plot allows to combine these aspects in one picture. The histogram in Figure 2.8 confirms that the modes have about the same height.

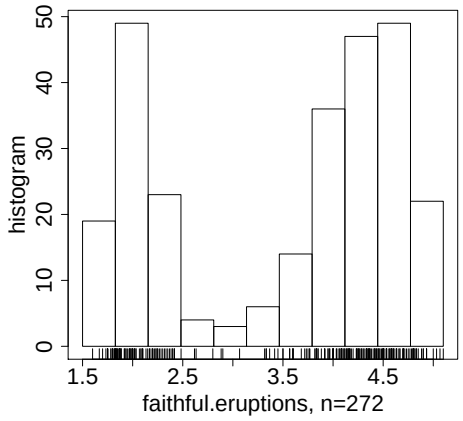

Figure 2.8: Histogram with 11 bins of the eruption durations of the Old Faithful geyser.

2.3.3 Melbourne Temperature Data

In Hyndman et al. (1996) the bifurcation to bimodality in the Melbourne temperature data set is pointed out. We use an extended version of the data set (Melbourne temperature data 1955-2007, provided by the Bureau of Meteorology, Victorian Climate Services Centre, Melbourne) and analyze the day by day difference in temperature at 3:00 pm, conditioned on today's temperature. The shorth plot in Figure 2.9 clearly indicates bimodality (and some skewness) when conditioning on high temperatures, and unimodality when conditioning on the lower temperatures.

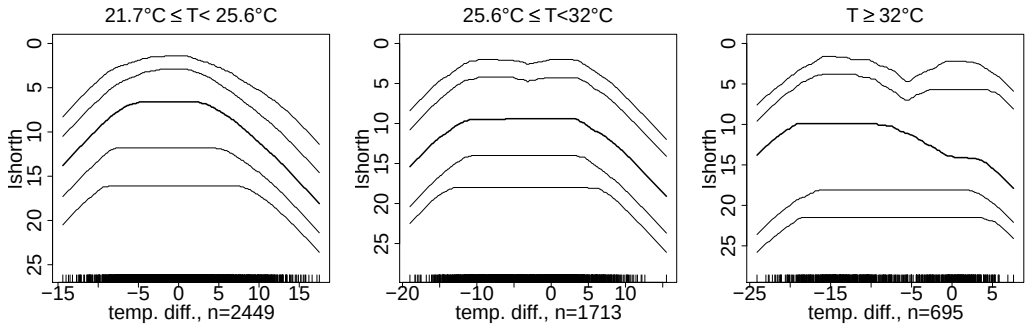

Figure 2.9: Melbourne day by day temperature difference at 3:00 pm conditioned on today's temperature; coverage levels (from top to bottom) $\alpha = 0.125, 0.25, 0.5, 0.75$ and 0.875.

2.3.4 Family Incomes in the UK

This example was used in Chaudhuri and Marron (1999) to point out the difficulty of selecting the "optimal" bandwidth in kernel density estimation. The data are $n = 7203$ net family incomes (rescaled to mean one) from the Family Expenditure Survey in the United Kingdom for the year 1975. For detailed discussion and analysis of the data see Schmitz and Marron (1992). In comparison to parametric approaches which typically lead to a unimodal representation of the data, in Schmitz and Marron (1992) the bimodality of the data is revealed by applying kernel density estimation. The bimodality can be explained by the different densities of pensioner households and non-pensioner households. The SiZer approach confirms this bimodality.

With these data, we want to compare the shorth plot and the SiZer map, see Figure 2.10. For the latter one, applied to the family income data, the reader is referred to Chaudhuri and Marron (1999) for a more detailed discussion. There also a comparison to a plot of kernel density estimators with various bandwidths can be found.

The first manifest characteristic revealed by the shorth plot is the right-skewness of the data. The top two lines ($\alpha = 0.0625,\ 0.125$) show the bimodality of the data: there is a high but short first mode and a nearly as high but broader second mode. The third coverage level ($\alpha = 0.25$) underlines this difference of mass concentration.

Some features are also revealed by the SiZer map. (Recall that the color scheme is blue (red) in locations where the curve is significantly increasing (decreasing), purple is used for a zero-derivative and grey indicates regions where the data are too sparse to make statements about significance.) Right skewness and one clear mode are apparent at all bandwidths, whereas the second mode can only be detected when looking at the "good" bandwidths. (As an indication, the "optimal" bandwidth given by Silverman's (1986) rule of thumb is marked with a white line in the SiZer map.) The SiZer map, however, offers no information about the height of the modes.

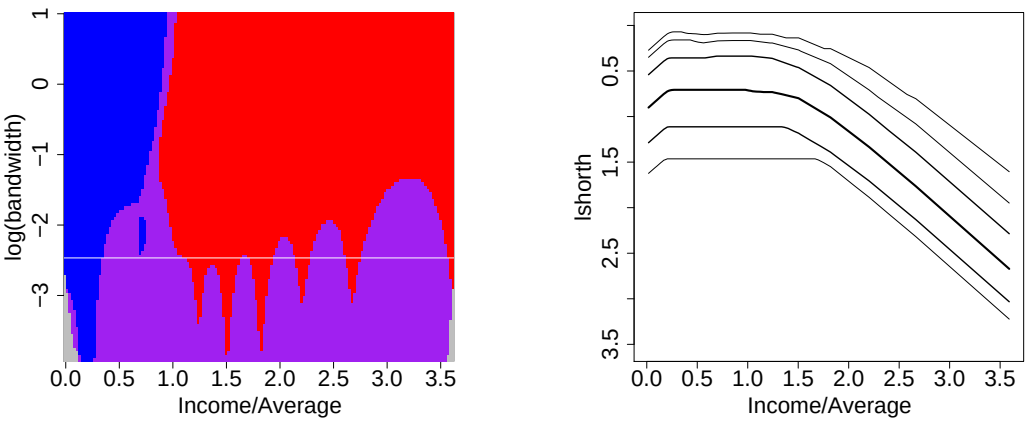

Figure 2.10: Family Expenditure Survey 1975: rescaled net household income. (left) SiZer map; "optimal" bandwidth $h = 0.0943$ marked by a white line. (right) Shorth plot with coverage levels (from top to bottom) $\alpha = 0.0625$, 0.125, 0.25, 0.5, 0.75 and 0.875.

2.4 Discussion

The α-shorth is a well-defined concept. Its length is studied in detail in Grübel (1988). The length of the shorth can be extended to higher dimensions by replacing the class of intervals by a class of sets (e.g., all ellipsoids) and length by "volume"; see Einmahl and Mason (1992) for the asymptotic behavior of these minimal volumes. In higher dimensions, however, there is no canonical class of sets, like the intervals in dimension one.

The shorth plot was introduced in Sawitzki (1994), but no further analysis or theory was provided. The asymptotic theory for the shorth plot, with convergence rate $n^{-\frac{1}{2}}$, can be found in Einmahl et al. (2010). A closely related idea is the balloonogram in Tukey and Tukey (1981). That multivariate procedure reduces in dimension one to considering the shortest interval, centered at a data point, that contains a certain number of data. In contrast to the balloonogram, the shorth plot avoids centering, thus reducing random fluctuation. No theory is provided, however, and also only one coverage level is used at a time.

The shorth plot is based on the concept of mass concentration, an idea shared with the excess density plot and the silhouette plot (Müller and Sawitzki, 1991). Excess density and silhouette plots are designed to detect the modes of a density. They use a global approach: there is no localization in x, like in the shorth plot. In Hyndman (1996), so-called highest density regions boxplots are introduced. These boxplots use mass concentration in a regression context.

Kernel density estimators with varying bandwidths are widely studied and somewhat related to our approach. The coverage α of the shorth plot bears some similarity with the bandwidth chosen for kernel estimation. The SiZer (Chaudhuri and Marron, 1999) is a kernel-based approach which studies simultaneously a wide range of bandwidths. Another approach that combines kernel estimation explicitly with detecting modes is that of the mode trees (Minnotte and Scott, 1993). Here the mode locations are plotted against the bandwidth of the density estimator with those modes. In the recent paper Dümbgen and Walther (2008), increases and decreases of a density are investigated through multiscale test statistics. Mass concentration is a local concept, but not, like a density, an infinitesimal concept. Therefore the shorth plot avoids the smoothing step and can be based directly on the empirical measure.

2.5 Supplementary Materials

R-package for the shorth plot: R-package "lshorth" containing code to compute the shorth plot. (zipped tar.gz file)

Meuse discharges (Section 2.3.1): Data set with the annual maximum discharges of the Meuse river from 1911 through 1995 at Borgharen in The Netherlands. (txt file)

Old Faithful geyser (Section 2.3.2): Data set with the eruption durations of the Old Faithful geyser in Yellowstone Park. (Data set "faithful$eruptions" available in R.)

Melbourne temperature (Section 2.3.3): Melbourne temperature 1955-2007, daily measurements. The data are part of the R-package "lshorth" and can be accessed via `http://lshorth.r-forge.r-project.org/data/melbourne/`.

Family Incomes (Section 2.3.4): Net family incomes from the Family Expenditure Survey in the United Kingdom for the year 1975. The data can be requested from the ESCR Data Archive at the University of Essex, `http://www.data-archive.ac.uk/ findingData/snDescription.asp?sn=3052`.

Chapter 3

Asymptotics of the shorth plot

[Based on joint work with J.H.J. Einmahl and G. Sawitzki, *Asymptotics of the shorth plot*, Journal of Statistical Planning and Inference (2010),140, 3003–3012.]

Abstract: The shorth plot is a nonparametric method to visualize probability mass concentration. It is based on the length of the shortest interval containing a certain fraction of the probability distribution and a point x. We establish functional central limit theorems (convergence rate $1/\sqrt{n}$) for the empirical shorth plot under natural conditions. The limiting process is not necessarily Gaussian. In the proofs, we generalize the Vervaat (1972) lemma to a collection of functions.

Keywords: Distribution diagnostics, functional central limit theorem, graphical methods, probability mass concentration.

3.1 Introduction

Graphical methods are of great importance in statistics. For detecting specific features of probability distributions, special tools have been developed, such as the PP- and QQ-plot for comparison with a given model distribution, or the box-and-whisker plot for a global view on location and scale. When

studying the modality of a distribution, silhouette plots and excess-mass plots can be used, see Müller and Sawitzki (1991). To obtain a general overview of a distribution, usually density estimators are employed. But the bandwidth choice may critically influence the qualitative features revealed or suggested by density estimators. Too large bandwidths lead to oversmoothing, too small bandwidths create artefacts. Plotting density estimates for different bandwidths simultaneously is difficult because the estimated densities overlap.

The *shorth plot* is a new, alternative method for depicting probability distributions, with competitive statistical behavior; see Einmahl et al. (2010) for a thorough study of the shorth plot from a data analytical point of view. In contrast to density estimation, for the shorth plot it is possible and recommended to plot several coverage levels of the shorth in one plot. Due to the monotonicity property of the shorth (see Section 3.2), the various plots do not overlap and display different qualitative characteristics of the distribution.

It is the aim of this paper to present the asymptotic behavior of the shorth plot. The shorth is the shortest interval containing a certain fraction of the distribution, originally half of the distribution, see Andrews et al. (1972). The length of the shorth can be a useful functional, as pointed out in Grübel (1988). There the asymptotic properties of the length of the shorth are studied and a convergence rate of $n^{-\frac{1}{2}}$ with a Gaussian limit is obtained. The critical conditions for Gaussianity are that the shorth interval is sufficiently pronounced, essentially this means that the shorth interval must not be in a flat part of the density, see Section 3.3 in Grübel (1988). In Einmahl and Mason (1992) it was shown that the good convergence of $n^{-\frac{1}{2}}$ is retained under much weaker conditions, including flat-part densities, however, the limit there can be non-Gaussian. For the shorth plot the length of the shorth is considered under a localization in $x \in \mathbb{R}$. This localization permits its use as a diagnostic tool. The convergence rate $n^{-\frac{1}{2}}$ is shown to hold uniformly in both x and the coverage level α. Recall that density estimators have a slower convergence rate. We present two functional central limit theorems, the first one - under natural and easily verifiable conditions - in which α is bounded away from 0 and 1. The second theorem extends this result to all coverage

levels $\alpha \in (0,1)$, under stronger conditions. For the proof, we derive a generalization to a collection of functions of the well-known Vervaat (1972) lemma, which might be of independent interest. For applications and examples of the shorth plot and comparison to other methods, we refer to Einmahl et al. (2010).

The paper is organized as follows. In the next section, precise definitions of the length of the shorth and the shorth plot are given, along with the most important properties. In Section 3.3 the main asymptotic results are presented. The proofs are deferred to Section 3.4.

3.2 The Shorth Plot

Let $X_1, \ldots, X_n$, $n \geq 1$, be i.i.d. random variables with common distribution function F and a corresponding probability measure P. The empirical measure P_n on the Borel sets $\mathbb{B}$ on $\mathbb{R}$ is defined by

$$P_n(B) = \frac{1}{n} \sum_{i=1}^{n} 1_B(X_i), \quad B \in \mathbb{B},$$

with 1_B denoting the indicator function. Denote with $\mathbb{I} = \{[a,b] : -\infty < a < b < \infty\}$ the class of closed intervals and with $\mathbb{I}_x = \{[a,b] : -\infty < a < b < \infty,\ x \in [a,b]\}$ the class of closed intervals containing $x \in \mathbb{R}$. Let $|\cdot|$ denote Lebesgue measure.

Definition 3.1. The length of the shorth at point $x \in \mathbb{R}$ for coverage level $\alpha \in (0,1)$ is

$$S_x(\alpha) = \inf\{|I| : P(I) \geq \alpha, I \in \mathbb{I}_x\}.$$

The definition in terms of a theoretical probability P has an immediate empirical counterpart, the empirical length of the shorth

$$S_{n,x}(\alpha) = \inf\{|I| : P_n(I) \geq \alpha, I \in \mathbb{I}_x\}.$$

We assume that the distribution function F is absolutely continuous with uniformly continuous density f with support $\mathcal{S} = (x_*, x^*)$, $-\infty \leq x_* < x^* \leq \infty$. Observe that F is strictly increasing on $\mathcal{S}$ and that f is bounded. $S_x(\alpha)$ exhibits the following properties.

- *Minimizing intervals:* For every α and x, there exists an interval I with length $S_x(\alpha)$ such that $x \in I$ and $P(I) = \alpha$.

- *Continuity:* For all α, $|S_x(\alpha) - S_y(\alpha)| \leq |x - y|$. Moreover, the function

$$(x, \alpha) \mapsto S_x(\alpha)$$

is continuous as a function of two variables.

- *Monotonicity:* For all x,

$$\alpha \mapsto S_x(\alpha)$$

is strictly increasing in α.

- *Invariance:* For all α,

$$x \mapsto S_x(\alpha)$$

is invariant under shift transformations and equivariant under scale transformations, that is when we apply a transformation $u' = cu + d$ (for some constants $c > 0, d$), then the new $S'_{x'}(\alpha)$ satisfies

$$S'_{x'}(\alpha) = cS_x(\alpha),$$

with $x' = cx + d$.

Definition 3.2. [Sawitzki, (1994)] The shorth plot is the graph of the function

$$x \mapsto S_x(\alpha),\ x \in \mathbb{R},$$

for (all or) a selection of coverages α.

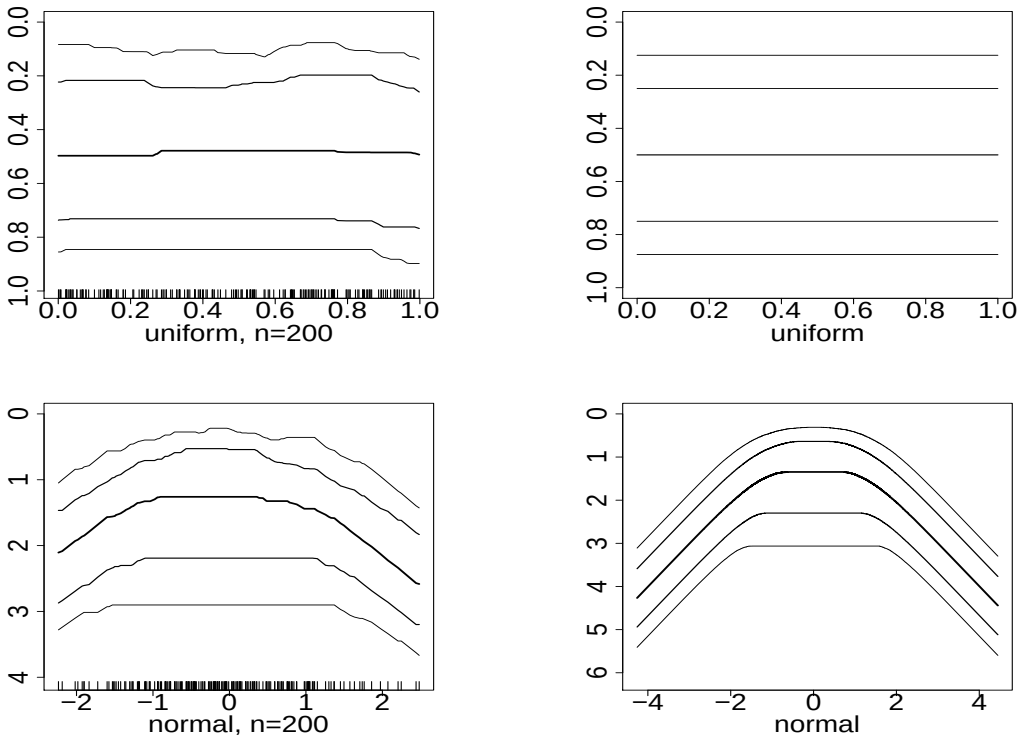

Figure 3.1: Empirical shorth plots for a standard uniform and a standard normal distribution based on a sample of size 200 and the corresponding theoretical shorth plots, for coverage levels (from top to bottom) $\alpha = 0.125$, 0.25, 0.5, 0.75, 0.875. Note that the vertical axis has a downward orientation.

The empirical shorth plot is the graph of

$$x \mapsto S_{n,x}(\alpha), \ x \in \mathbb{R}.$$

Figure 3.1 shows the empirical shorth plots for samples of size 200 of a standard uniform and a standard normal distribution and the theoretical shorth plots. Several coverage levels α are chosen to display various features of the data simultaneously. Following Einmahl et al. (2010), the "dyadic" scale 0.125, 0.25, 0.5, 0.75, 0.875 is used for the α-values. Due to the Monotonicity property (Section 3.2), the multiple scales can be displayed in one figure without overlaps. The local behavior of the distribution, in particular near modes, is

given by small coverage levels, whereas high coverage levels show the range of the distribution. Intermediate levels inform about the skewness of the overall distribution shape.

3.3 Main Results

In this section we present the main results on the asymptotic behavior of the empirical shorth plot. Let $\mathbb{I}^* = \mathbb{I} \cup \{\mathbb{R}, \emptyset\}$. Define for each $n \geq 1$ the empirical process indexed by intervals to be

$$U_n(I) = n^{\frac{1}{2}}\{P_n(I) - P(I)\}, \quad I \in \mathbb{I}^*.$$

Introduce the pseudometric d_0 defined on $\mathbb{B}$ by

$$d_0(B_1, B_2) = P(B_1 \triangle B_2), \quad \text{for } B_1, B_2 \in \mathbb{B},$$

with $B_1 \triangle B_2 = (B_1 \setminus B_2) \cup (B_2 \setminus B_1)$. Let B_P be a bounded, mean zero Gaussian process indexed by $\mathbb{I}^*$, uniformly continuous in d_0, with covariance function $P(A_1 \cap A_2) - P(A_1)P(A_2)$, $A_1, A_2 \in \mathbb{I}^*$. Then, by the functional central limit theorem and the Skorohod representation theorem (see Shorack and Wellner (1986), p. 110 and p. 93, respectively), there exist $\tilde{B}_P \stackrel{d}{=} B_P$ and a sequence $\tilde{U}_n \stackrel{d}{=} U_n$ such that

$$\sup\{|\tilde{U}_n(I) - \tilde{B}_P(I)| : I \in \mathbb{I}^*\} \to 0 \quad \text{a.s. as } n \to \infty. \tag{3.1}$$

Henceforth we will drop the tildes from the notation.

We will need the following assumption:

(A) There exist $x_1, x_2 \in [x_*, x^*]$, $x_1 \leq x_2$, such that f is strictly increasing on $(x_*, x_1]$, constant on $[x_1, x_2]$, and strictly decreasing on $[x_2, x^*)$; also $\lim_{x \downarrow x_*} f(x) = \lim_{x \uparrow x^*} f(x)$.

We introduce some more notation. Write $g_x = \frac{1}{\frac{d}{d\alpha}S_x(\alpha)}$; see Section 3.4, Lemma 3.3 for the existence of g_x. Set $T_{x,0} = \{\emptyset\}$, $T_{x,1} = \{\mathbb{R}\}$, $x \in \mathbb{R}$ and

$$T_{x,\alpha} = \{I \in \mathbb{I}_x : \ |I| = S_x(\alpha),\ P(I) = \alpha\} \quad \text{for } \ 0 < \alpha < 1,\ x \in \mathbb{R}.$$

For any $x \in \mathbb{R}$ let

$$B_x(\alpha) = \sup\{B_P(I) : \ I \in T_{x,\alpha}\}, \quad 0 \le \alpha \le 1.$$

The process $B_x(\alpha)$, $x \in \mathbb{R}$, $\alpha \in (0,1)$ is the candidate limit; observe that only the shortest intervals, i.e. those in $T_{x,\alpha}$, determine this process.

Consider the *shorth plot process*:

$$Q_{n,x}(\alpha) = g_x(\alpha) n^{\frac{1}{2}} \left(S_{n,x}(\alpha) - S_x(\alpha)\right), \quad 0 < \alpha < 1, \ \ x \in \mathbb{R}.$$

Theorem 3.1. *Under the assumptions on F of Section 3.2 and assumption (A) we have for all $0 < \eta < 1/2$, on the probability space of (3.1),*

$$\sup_{x \in \mathbb{R}} \sup_{\eta \le \alpha \le 1-\eta} |Q_{n,x}(\alpha) + B_x(\alpha)| \to 0 \quad a.s. \quad as \ \ n \to \infty. \tag{3.2}$$

We need two additional assumptions to extend the convergence on $[\eta, 1-\eta]$ in (3.2) to convergence on the entire interval $(0,1)$. The first one is the classical Csörgő and Révész (1978) condition:

(B) If $\lim_{x \downarrow x_*} f(x) = 0$, then f' exists on $\mathcal{S}$ and for some $0 < M < \infty$

$$\sup_{x \in \mathcal{S}} F(x)\left(1 - F(x)\right) \frac{|f'(x)|}{f^2(x)} < M.$$

For the second assumption, let $I_\alpha = [a, b]$ be a shortest interval such that $P(I_\alpha) = \alpha$, $\alpha \in (0,1)$; note that $f(a) = f(b)$. If $\lim_{x \downarrow x_*} f(x) = 0$, then for large enough α, I_α is unique. For such an α define $\lambda_\alpha = F(a)/(1-\alpha)$.

(C) If $\lim_{x \downarrow x_*} f(x) = 0$, then $0 < \lim_{\alpha \uparrow 1} \lambda_\alpha < 1$.

Theorem 3.2. *In addition to the assumptions of Theorem 3.1, suppose (B) and (C) hold. Then on the probability space of (3.1),*

$$\sup_{x \in \mathbb{R}} \sup_{0<\alpha<1} |Q_{n,x}(\alpha) + B_x(\alpha)| \xrightarrow{\mathbb{P}} 0 \quad as \;\; n \to \infty.$$

Remark 3.1. Note that if $T_{x,\alpha}$ contains at least two sets I_1 and I_2 with $P(I_1 \triangle I_2) > 0$, then $B_x(\alpha)$ is not a normal random variable and $\mathbb{E}B_x(\alpha) > 0$. Since B_P is bounded, $B_\cdot(\cdot)$ is a bounded process on $\mathbb{R} \times [0,1]$ with $B_x(0) = B_x(1) = 0$ almost surely. In Section 3.4 we show that $(B_x)_{x \in \mathbb{R}}$ is a collection of uniformly equicontinuous functions on $[0,1]$.

Next we consider some general examples of the limiting process $B_x(\alpha)$, $x \in \mathbb{R}$, $0 < \alpha < 1$. Denote with $\mathcal{F}$ the class of distribution functions satisfying the assumptions of Theorem 3.1 with $x_1 = x_2$ in (A), and for $F \in \mathcal{F}$ write

$$\mathbb{I}_F = \{[F(z_1), F(z_2)] : \, f(z_1) = f(z_2), \, x_* < z_1 < z_2 < x^*\}.$$

For a given $F_0 \in \mathcal{F}$, define $\mathcal{F}_{F_0} = \{F \in \mathcal{F} : \mathbb{I}_F = \mathbb{I}_{F_0}\}$. Now write

$$\bar{B}(u, \alpha) = B_{F^{-1}(u)}(\alpha), \;\; 0 < u, \alpha < 1.$$

We will actually consider $\bar{B}$ instead of $B_x(\alpha)$, $x \in \mathbb{R}$, $0 < \alpha < 1$. We have that all $F \in \mathcal{F}_{F_0}$ lead to the same process $\bar{B}$ and this process is Gaussian. In particular, all symmetric distributions lead to the same $\bar{B}$. If we write

$$B(u) = B_P\left(\left(-\infty, F^{-1}(u)\right]\right), \;\; 0 < u < 1,$$

for the 'underlying' standard Brownian bridge (see (3.1)), we obtain for all the symmetric distributions

$$\bar{B}(u, \alpha) = \begin{cases} B(u + \alpha) - B(u), & \text{for } u < \frac{1}{2} - \frac{1}{2}\alpha \\ B(\frac{1}{2} + \frac{1}{2}\alpha) - B(\frac{1}{2} - \frac{1}{2}\alpha), & \text{for } \frac{1}{2} - \frac{1}{2}\alpha \le u \le \frac{1}{2} + \frac{1}{2}\alpha \\ B(u) - B(u - \alpha), & \text{for } u > \frac{1}{2} + \frac{1}{2}\alpha. \end{cases}$$

Remark 3.2. It readily follows that for every interval $I \in \mathbb{I}$, there exists an $x \in \mathbb{R}$ such that $I \in T_{x,P(I)}$.

This implies that

$$\sup_{x\in\mathbb{R}} \sup_{0\le\alpha\le 1} |B_x(\alpha)| \;=\; \sup_{I\in\mathbb{I}} |B_P(I)| \stackrel{d}{=} \sup_{0\le\alpha,\beta\le 1} |B(\alpha) - B(\beta)|, \qquad (3.3)$$

with B a standard Brownian bridge. The right-hand side of (3.3) is the limiting distribution of the Kuiper statistic, see e.g. Shorack and Wellner (1986), p. 144.

3.4 Proofs

3.4.1 Proof of Theorem 3.1

The proof of Theorem 3.1 is based on a number of lemmas and a proposition, which use assumption (A). In Lemmas 3.1 and 3.2 we present certain extensions of the conditions (C_6) and (C_8) in Einmahl and Mason (1992), respectively.

Lemma 3.1. *For every $\varepsilon > 0$, whenever $x \in \mathbb{R}$, $0 \le \alpha_1, \alpha_2 \le 1$ with $|\alpha_1 - \alpha_2| < \varepsilon$ and $I_1 \in T_{x,\alpha_1}$, there is an $I_2 \in T_{x,\alpha_2}$ with $d_0(I_1, I_2) < \varepsilon$.*

Proof Write $\beta = P([x_1, x_2])$, x_1, x_2 from assumption (A). Observe that $T_{x,\alpha}$ contains infinitely many intervals if and only if $x \in (x_1, x_2)$ and $\alpha < \beta$, otherwise it contains exactly one interval. From this it follows that for $\alpha_1 < \alpha_2$, an I_2 as in the lemma can be found with $I_2 \supset I_1$. Similarly, for $\alpha_1 > \alpha_2$, we can take $I_2 \subset I_1$. □

Lemma 3.2. *For every $\varepsilon > 0$ there exists a $\delta > 0$ such that whenever $I \in \mathbb{I}_x$, $x \in \mathbb{R}$, satisfies $0 < \alpha - \delta < P(I) < \alpha < 1$ and $|I| < S_x(\alpha)$, there is an $I' \in T_{x,S_x^{-1}(|I|)}$ such that $d_0(I, I') < \varepsilon$.*

Proof The proof can be given along the same lines as in Example 2 in Einmahl and Mason (1992). We will omit details. □

Let J be an open or closed interval and consider a function $f : J \to \mathbb{R}$; let $\delta > 0$. The modulus of continuity of f is defined by

$$\omega(f, \delta) = \sup\{|f(u) - f(v)| : u, v \in J, |u - v| \leq \delta\}. \tag{3.4}$$

Lemma 3.3. *For all $x \in \mathbb{R}$, g_x exists on $(0, 1)$ and is positive. Moreover, $(g_x)_{x\in\mathbb{R}}$ is a collection of uniformly equicontinuous functions on $(0, 1)$, i.e.*

$$\lim_{\delta\downarrow 0} \sup_{x\in\mathbb{R}} \omega(g_x, \delta) = 0.$$

In addition, for any $0 < \varepsilon < \frac{1}{2}$,

$$\inf_{x\in\mathbb{R}} \inf_{\varepsilon\leq\alpha\leq 1-\varepsilon} g_x(\alpha) > 0.$$

Proof Let $I_{x,\alpha} \in T_{x,\alpha}$. If x is on the boundary of $I_{x,\alpha}$, then $g_x(\alpha) = f(y)$, where y is the other endpoint of the interval $I_{x,\alpha}$. (Let, for instance, $x < y$. Then $S_x(\alpha) = F^{-1}(F(x) + \alpha) - x$, therefore $\frac{\partial}{\partial\alpha} S_x(\alpha) = \frac{1}{f(F^{-1}(F(x)+\alpha))} = \frac{1}{f(y)}$, and hence $g_x(\alpha) = f(y)$.) If x is not on the boundary of $I_{x,\alpha} =: [y_1, y_2]$, then, since $\lim_{x\downarrow x_*} f(x) = \lim_{x\uparrow x_*} f(x)$, $g_x(\alpha) = f(y_1) = f(y_2)$. Hence we obtain

$$\sup_{x\in\mathbb{R}} \sup_{\substack{|\alpha-\beta|\leq\delta \\ 0<\alpha,\beta<1}} |g_x(\alpha) - g_x(\beta)| \leq \sup_{\substack{|\alpha'-\beta'|\leq\delta \\ 0<\alpha',\beta'<1}} \left|f(F^{-1}(\alpha')) - f(F^{-1}(\beta'))\right|.$$

Now the uniform continuity of $f \circ F^{-1}$ on $(0, 1)$ yields the uniform equicontinuity of $(g_x)_{x\in\mathbb{R}}$.

Let $0 < \varepsilon < \frac{1}{2}$ and $a < b$ such that $P([a, b]) = 1 - \varepsilon$ and $f(a) = f(b) > 0$. Then it follows that

$$\inf_{x\in\mathbb{R}} \inf_{\varepsilon\leq\alpha\leq 1-\varepsilon} g_x(\alpha) \geq f(a). \qquad \square$$

Set

$$\bar{P}_{n,x}(\alpha) = \sup\{P_n(I) : \ |I| \le S_x(\alpha), \ I \in \mathbb{I}_x\}, \quad 0 < \alpha < 1, \ \ x \in \mathbb{R},$$

$\bar{P}_{n,x}(0) = 0$ and $\bar{P}_{n,x}(1) = 1$. Consider the process

$$\bar{U}_{n,x}(\alpha) = n^{\frac{1}{2}}(\bar{P}_{n,x}(\alpha) - \alpha), \quad 0 \le \alpha \le 1, \ \ x \in \mathbb{R}.$$

This is the right place to describe the main steps in the proof of Theorem 3.1. First we present in Proposition 3.1 the appropriate convergence result for the "uniformized" process $\bar{U}_{n,x}$. Next we "invert" this statement to get a similar convergence result for a uniformized quantile-type process (Corollary 3.1), and finally we obtain the theorem by "stretching out" this process in the vertical direction.

Proposition 3.1. *On the probability space of (3.1),*

$$\sup_{x\in\mathbb{R}} \sup_{0\le\alpha\le1} |\bar{U}_{n,x}(\alpha) - B_x(\alpha)| \to 0 \quad a.s. \ as \ n \to \infty. \tag{3.5}$$

Proof The proof follows closely the lines of that in Proposition 3.1 in Einmahl and Mason (1992), but now the supremum over x has also to be taken into account. Clearly it is sufficient to show that $\sup_{x\in\mathbb{R}} \sup_{0<\alpha<1} |\bar{U}_{n,x}(\alpha) - B_x(\alpha)| \to 0$ almost surely as $n \to \infty$. First we show that

$$\limsup_{n\to\infty} \sup_{x\in\mathbb{R}} \sup_{0<\alpha<1} (B_x(\alpha) - \bar{U}_{n,x}(\alpha)) \le 0 \quad \text{a.s.} \tag{3.6}$$

For any $0 < \alpha < 1$ and $x \in \mathbb{R}$,

$$\begin{aligned}
B_x(\alpha) - \bar{U}_{n,x}(\alpha) \ &\le \ \sup\{B_P(I) : I \in T_{x,\alpha}\} \\
&\qquad -n^{\frac{1}{2}} \left(\sup\{P_n(I) : |I| \le S_x(\alpha), \ P(I) = \alpha, \ I \in \mathbb{I}_x\} - \alpha\right) \\
&= \ \sup\{B_P(I) : I \in T_{x,\alpha}\} - n^{\frac{1}{2}} \sup\{P_n(I) - P(I) : I \in T_{x,\alpha}\} \\
&\le \ \sup\{B_P(I) - U_n(I) : I \in T_{x,\alpha}\} \\
&\le \ \sup\{B_P(I) - U_n(I) : I \in \mathbb{I}\}.
\end{aligned}$$

Now (3.6) follows from (3.1).

It remains to show that

$$\limsup_{n\to\infty} \sup_{x\in\mathbb{R}} \sup_{0<\alpha<1} \left(\bar{U}_{n,x}(\alpha) - B_x(\alpha)\right) \le 0 \quad \text{a.s.}$$

For any $0 < \alpha < 1$ and $x \in \mathbb{R}$ we have

$$\begin{aligned}
&\bar{U}_{n,x}(\alpha) - B_x(\alpha) \\
&\quad \le \left\{ n^{\frac{1}{2}} \left(\sup\{P_n(I) : |I| \le S_x(\alpha),\ \alpha - n^{-\frac{1}{4}} < P(I) \le \alpha,\ I \in \mathbb{I}_x\} - \alpha \right) \right. \\
&\qquad \left. - B_x(\alpha) \right\} \\
&\quad \vee \left\{ n^{\frac{1}{2}} \left(\sup\{P_n(I) : P(I) \le \alpha - n^{-\frac{1}{4}},\ I \in \mathbb{I}_x\} - \alpha \right) - B_x(\alpha) \right\}. \qquad (3.7)
\end{aligned}$$

The second term in the right-hand side of (3.7) is bounded from above by

$$\begin{aligned}
&\sup\left\{ n^{\frac{1}{2}} \left(P_n(I) - P(I) - n^{-\frac{1}{4}} \right) : I \in \mathbb{I}_x \right\} - B_x(\alpha) \\
&\quad \le \sup\{|U_n(I) - B_P(I)| : I \in \mathbb{I}_x\} + \sup\{|B_P(I)| : I \in \mathbb{I}_x\} \\
&\qquad\quad + \sup\{|B_P(I')| : I' \in T_{x,\alpha}\} - n^{\frac{1}{4}} \\
&\quad \le \sup\{|U_n(I) - B_P(I)| : I \in \mathbb{I}\} + 2\sup\{|B_P(I)| : I \in \mathbb{I}\} - n^{\frac{1}{4}},
\end{aligned}$$

which, by (3.1) and the boundedness of B_P, converges almost surely to $-\infty$, as $n \to \infty$.

Next consider the first term in the right-hand side of (3.7). For any $0 < \alpha < 1$ and $x \in \mathbb{R}$, this term is equal to

$$\begin{aligned}
&\sup\left\{ n^{\frac{1}{2}}(P_n(I) - \alpha) : |I| \le S_x(\alpha),\ \alpha - n^{-\frac{1}{4}} < P(I) \le \alpha,\ I \in \mathbb{I}_x \right\} - B_x(\alpha) \\
&\quad \le \sup\{|U_n(I) - B_P(I)| : I \in \mathbb{I}_x\} \\
&\qquad\quad + \sup\left\{ B_P(I) : |I| \le S_x(\alpha),\ \alpha - n^{-\frac{1}{4}} < P(I) \le \alpha,\ I \in \mathbb{I}_x \right\} - B_x(\alpha).
\end{aligned}$$

The first term tends to zero, uniformly in $x \in \mathbb{R}$, almost surely as $n \to \infty$ because of $\mathbb{I}_x \subset \mathbb{I}$ and (3.1), so the proof of (3.5) will be complete if we show

$$\begin{aligned}
\sup_{x\in\mathbb{R}} \sup_{0<\alpha<1} &\left\{ \sup\left\{ B_P(I) : |I| \le S_x(\alpha),\ \alpha - n^{-\frac{1}{4}} < P(I) \le \alpha,\ I \in \mathbb{I}_x \right\} \right. \\
&\left. - \sup\left\{ B_P(I') : I' \in T_{x,\alpha} \right\} \right\} \to 0 \ \text{ as } \ n \to \infty. \qquad (3.8)
\end{aligned}$$

By Lemma 3.2 combined with Lemma 3.1, and uniform continuity of B_P for any $\eta > 0$, we have for all large n

$$\sup_{x\in\mathbb{R}} \sup_{0<\alpha<1} \Big| \sup\Big\{B_P(I) : |I| \le S_x(\alpha),\ \alpha - n^{-\frac{1}{4}} < P(I) \le \alpha,\ I \in \mathbb{I}_x\Big\} - \sup\ \big\{B_P(I') : I' \in T_{x,\alpha}\big\}\Big| \ \le\ \eta.$$

Since $\eta > 0$ is arbitrary, this implies (3.8). □

For $s : [0,1] \to \mathbb{R}$, write $\|s\| = \sup\{|s(\alpha)| : 0 \le \alpha \le 1\}$. Let Id denote the identity function. Recall the definition of ω in (3.4).

Lemma 3.4. *Let* Γ *be a nondecreasing function on* $[0,1]$ *with* $\Gamma(0) = 0$ *and* $\Gamma(1) = 1$. *Define* $\Gamma^{-1}(\alpha) = \inf\{\beta : \Gamma(\beta) \ge \alpha\}$, $0 \le \alpha \le 1$. *Then*

$$\|\Gamma + \Gamma^{-1} - 2Id\| \le \omega(\Gamma - Id,\ \|\Gamma - Id\|).$$

Proof Write $S = [-\|\Gamma - Id\|,\ \|\Gamma - Id\|]$. We have

$$\begin{aligned} \Gamma^{-1}(\alpha) - \alpha &= \inf\{\beta - \alpha : \Gamma(\beta) \ge \alpha\} \\ &= \inf\{\beta - \alpha \in S : \Gamma(\beta) \ge \alpha\}. \end{aligned} \tag{3.9}$$

The second equality in (3.9) follows, since for $\beta - \alpha > \|\Gamma - Id\|$ we have $\alpha - \beta < -\|\Gamma - Id\| \le \Gamma(\beta) - \beta$ and hence $\Gamma(\beta) > \alpha$ (therefore $\inf\{\beta - \alpha \in (\|\Gamma - Id\|, \infty) : \Gamma(\beta) \ge \alpha\} = \|\Gamma - Id\|$); for $\beta - \alpha < -\|\Gamma - Id\|$, we have $\alpha - \beta > \|\Gamma - Id\|$ and therefore $\alpha - \beta \le \Gamma(\beta) - \beta$ or $\alpha \le \Gamma(\beta)$ is impossible. Thus

$$\begin{aligned} \Gamma^{-1}(\alpha) - \alpha &= \inf\{\beta - \alpha \in S : \beta - \alpha \ge \alpha - \Gamma(\alpha) + \beta - \Gamma(\beta) - (\alpha - \Gamma(\alpha))\} \\ &\ge \inf\{\beta - \alpha \in S : \beta - \alpha \ge \alpha - \Gamma(\alpha) - \omega(\Gamma - Id, \|\Gamma - Id\|)\} \\ &\ge \alpha - \Gamma(\alpha) - \omega(\Gamma - Id, \|\Gamma - Id\|). \end{aligned}$$

Similarly,

$$\begin{aligned}\Gamma^{-1}(\alpha)-\alpha &\leq \inf\{\beta-\alpha\in S:\beta-\alpha\geq\alpha-\Gamma(\alpha)+\omega(\Gamma-Id,\|\Gamma-Id\|)\}\\ &\leq \alpha-\Gamma(\alpha)+\omega(\Gamma-Id,\|\Gamma-Id\|),\end{aligned}$$

and hence

$$-\omega(\Gamma-Id,\|\Gamma-Id\|)\ \leq\ \Gamma^{-1}(\alpha)-2\alpha+\Gamma(\alpha)\ \leq\ \omega(\Gamma-Id,\|\Gamma-Id\|).\square$$

We will use this lemma to establish a generalization to a collection of functions of the well-known lemma in Vervaat (1972).

Lemma 3.5. *Let $\Gamma_{n,x}$ be a collection of nondecreasing functions on $[0,1]$ indexed by $n\in\mathbb{N}$ and $x\in\mathfrak{X}$ (some index set). Assume for all n and x, $\Gamma_{n,x}(0)=0$ and $\Gamma_{n,x}(1)=1$. Moreover, let b_x, $x\in\mathfrak{X}$, be a collection of uniformly bounded ($\sup_{x\in\mathfrak{X}}\sup_{0\leq\alpha\leq1}|b_x(\alpha)|<\infty$) and uniformly equicontinuous functions on $[0,1]$. Finally let $(m_n)_{n\in\mathbb{N}}$ be a sequence of positive numbers tending to infinity.*
If, as $n\to\infty$,

$$\sup_{x\in\mathfrak{X}}\sup_{0\leq\alpha\leq1}|m_n(\Gamma_{n,x}(\alpha)\ -\ \alpha)-b_x(\alpha)|\to0,$$

then

$$\sup_{x\in\mathfrak{X}}\sup_{0\leq\alpha\leq1}|m_n(\Gamma_{n,x}^{-1}(\alpha)\ -\ \alpha)+b_x(\alpha)|\to0.$$

Proof If $s_x:[0,1]\to\mathbb{R}$ for every $x\in\mathfrak{X}$, write $|||s_x|||=\sup\{|s_x(\alpha)|:\ 0\leq\alpha\leq1,\ x\in\mathfrak{X}\}$. Define $D_n=|||m_n(\Gamma_{n,x}-Id)-b_x|||$. From Lemma 3.4 we have

$$\begin{aligned}&|||m_n(\Gamma_{n,x}^{-1}-Id)\ +\ b_x|||\\ &\quad\leq|||m_n(\Gamma_{n,x}^{-1}-Id+\Gamma_{n,x}-Id)|||+\ |||-\left[m_n(\Gamma_{n,x}-Id)-b_x\right]|||\\ &\quad\leq\ m_n\sup_{x\in\mathfrak{X}}\omega(\Gamma_{n,x}-Id,\|\Gamma_{n,x}-Id\|)\ +\ D_n\end{aligned}$$

$$
\begin{aligned}
&\leq \sup_{x\in\mathfrak{X}} \Bigg\{ \sup_{|\beta-\alpha|\leq\|\Gamma_{n,x}-Id\|} |m_n(\Gamma_{n,x}(\alpha)-\alpha) - b_x(\alpha) - m_n(\Gamma_{n,x}(\beta)-\beta) + b_x(\beta)| \\
&\quad + \sup_{|\beta-\alpha|\leq\|\Gamma_{n,x}-Id\|} |b_x(\alpha)-b_x(\beta)| \Bigg\} + D_n \\
&\leq \sup_{x\in\mathfrak{X}} \Bigg\{ 2 \sup_{0\leq\alpha\leq 1} |m_n(\Gamma_{n,x}(\alpha)-\alpha) - b_x(\alpha)| \\
&\quad + \sup_{|\beta-\alpha|\leq\|\Gamma_{n,x}-Id\|} |b_x(\beta)-b_x(\alpha)| \Bigg\} \; + \; D_n \\
&\leq \sup_{x\in\mathfrak{X}} \omega\left(b_x\,,\, \|\Gamma_{n,x}-Id\|\right) \; + \; 3D_n \\
&\leq \sup_{x\in\mathfrak{X}} \omega\left(b_x\,,\, \frac{1}{m_n} \sup_{0\leq\alpha\leq 1} |m_n(\Gamma_{n,x}(\alpha)-\alpha) - b_x(\alpha)| + \sup_{0\leq\alpha\leq 1} \frac{|b_x|}{m_n}\right) \; + \; 3D_n \\
&\leq \sup_{x\in\mathfrak{X}} \omega\left(b_x\,,\, \frac{D_n + \|\!|b_x|\!\|}{m_n}\right) + 3D_n \to 0 \;\text{ as }\; n\to\infty\,. \qquad \square
\end{aligned}
$$

Define

$$V_{n,x}(\beta) = \inf\{\alpha : \bar{P}_{n,x}(\alpha) \geq \beta,\ 0\leq\alpha\leq 1\}, \quad 0\leq\beta\leq 1,\ \ x\in\mathbb{R},$$

$$\bar{Q}_{n,x}(\alpha) = n^{\frac{1}{2}}(V_{n,x}(\alpha)-\alpha), \quad 0\leq\alpha\leq 1,\ \ x\in\mathbb{R}.$$

It is immediate from Lemma 3.1 and the continuity of B_P, that $(B_x)_{x\in\mathbb{R}}$ is a collection of uniformly equicontinuous functions on $[0,1]$. Hence combining Lemma 3.5 and Proposition 3.1, we obtain the following important result.

Corollary 3.1. *On the probability space of (3.1), as $n\to\infty$,*

$$\sup_{x\in\mathbb{R}} \sup_{0\leq\alpha\leq 1} |\bar{Q}_{n,x}(\alpha) + B_x(\alpha)| \to 0 \;\; a.s.$$

and hence

$$\sup_{x\in\mathbb{R}} \sup_{0\leq\alpha\leq 1} |V_{n,x}(\alpha)-\alpha| \to 0 \;\; a.s.$$

Define $S_x(0) = \lim_{\alpha \downarrow 0} S_x(\alpha)$. Similar to Lemma 3.1 in Einmahl and Mason (1992) we can show:

Lemma 3.6. *With probability 1, for all* $0 < \alpha < 1$ *and* $x \in \mathbb{R}$,

$$S_{n,x}(\alpha) = S_x(V_{n,x}(\alpha)).$$

Proof of Theorem 3.1: For each $\eta \leq \alpha \leq 1-\eta$ and $x \in \mathbb{R}$ we get by Lemma 3.6 and the mean value theorem, that almost surely

$$\begin{aligned} Q_{n,x}(\alpha) + B_x(\alpha) &= g_x(\alpha)\, n^{\frac{1}{2}}\, (S_x(V_{n,x}(\alpha)) - S_x(\alpha)) + B_x(\alpha) \\ &= \frac{g_x(\alpha)}{g_x(\theta_{n,x})}\, \bar{Q}_{n,x}(\alpha) + B_x(\alpha) \\ &= \frac{g_x(\alpha)}{g_x(\theta_{n,x})} \left(\bar{Q}_{n,x}(\alpha) + B_x(\alpha)\right) - \frac{g_x(\alpha) - g_x(\theta_{n,x})}{g_x(\theta_{n,x})} B_x(\alpha), \end{aligned} \tag{3.10}$$

where $\theta_{n,x}$ lies between α and $V_{n,x}(\alpha)$. Assertion (3.2) follows from Corollary 3.1 and Lemma 3.3. □

3.4.2 Proof of Theorem 3.2

For the proof of Theorem 3.2 we need three more auxiliary results.

Fact 3.1. [Lemma 1 in Csörgő and Révész (1978)] Under the assumptions of Theorem 3.2, in particular condition (B), we have

$$\frac{f\left(F^{-1}(\alpha)\right)}{f\left(F^{-1}(\beta)\right)} \leq \left[\frac{\beta \vee \alpha}{\beta \wedge \alpha} \frac{1-(\beta \wedge \alpha)}{1-(\beta \vee \alpha)}\right]^M \quad \text{for all } \; 0 < \alpha, \beta < 1.$$

Fact 3.2. [Lemma 3.2 in Einmahl and Mason (1992)] Let $(Y_{n,k})_{n\geq 1,\, k\geq 1}$ be a double sequence of random variables such that for each $n, k \in \mathbb{N}$, $Y_{n,k}$ is Binomial$(n, 2^{-k})$. Then

$$Y_n := \sup_{k \in \mathbb{N}}\; n^{-1}\, 2^k\, Y_{n,k} = O_{\mathbb{P}}(1).$$

Lemma 3.7. *On the probability space of (3.1),*

$$\sup_{x\in\mathbb{R}} \sup_{0<\alpha<1} \frac{1-\alpha}{1-V_{n,x}(\alpha)} = O_{\mathbb{P}}(1).$$

Proof For $k \in \mathbb{N}$, $x \in \mathbb{R}$ choose $I_{k,x} \in T_{x,1-2^{-k}}$ and for $1-2^{-k} \le \alpha < 1-2^{-k-1}$ set $I_{\alpha,x} = I_{k,x}$. Following the proof of Lemma 3.3 in Einmahl and Mason (1992), we can now show that

$$\sup_{x\in\mathbb{R}} \sup_{0<\alpha<1} \frac{1-\alpha}{1-V_{n,x}(\alpha)} \;\le\; 2 \vee \left\{ \sup_{x\in\mathbb{R}} \sup_{k\ge 1} 2^{k+1}(1-P_n(I_{k,x})) \right\}. \tag{3.11}$$

Write $W_k := [F^{-1}(2^{-k}), F^{-1}(1-2^{-k})]$. Then $I_{k,x} \supset W_k$ holds for every $x \in \mathbb{R}$. Hence the second term of the right-hand side of (3.11) is bounded from above by

$$\sup_{k\ge 1} 2^{k+1}(1-P_n(W_k)). \tag{3.12}$$

Since $n(1-P_n(W_k))$ is Binomial$(n, 2^{-(k-1)})$, Fact 3.2 yields that the expression in (3.12) is $O_{\mathbb{P}}(1)$. □

Proof of Theorem 3.2: Theorem 3.1 states that for all $0 < \eta < 1/2$

$$\sup_{x\in\mathbb{R}} \sup_{\eta\le\alpha\le 1-\eta} |Q_{n,x}(\alpha) + B_x(\alpha)| \to 0 \quad \text{a.s. as } n \to \infty. \tag{3.13}$$

If $\lim_{x\downarrow x_*} f(x) > 0$, then $\inf_{x\in\mathbb{R}} \inf_{0<\alpha<1} g_x(\alpha) > 0$ and Theorem 3.2 holds using the same argument as in the proof of Theorem 3.1.

So in the sequel we assume $\lim_{x\downarrow x_*} f(x) = 0$. Because of (3.13) we only need to consider the supremum on the region where $x \in \mathbb{R}$ and $\alpha < \eta$ and on the region $x \in \mathbb{R}$ and $\alpha > 1-\eta$. We have from (3.10):

$$|Q_{n,x}(\alpha) + B_x(\alpha)| \le \left| \frac{g_x(\alpha)}{g_x(\theta_{n,x})} \left(\bar{Q}_{n,x}(\alpha) + B_x(\alpha) \right) \right| + \left| \left(\frac{g_x(\alpha)}{g_x(\theta_{n,x})} - 1 \right) B_x(\alpha) \right|.$$

Therefore it follows from a routine argument, using Corollary 3.1 and the equicontinuity of $(B_x)_{x\in\mathbb{R}}$ in conjunction with $B_x(0) = B_x(1) = 0$ for all $x \in \mathbb{R}$ almost surely, that it is sufficient to show that for small enough $\eta > 0$, $g_x(\alpha)/g_x(\theta_{n,x})$ is bounded in probability uniformly over both regions.

First we will show

$$\sup_{\alpha<\eta}\sup_{x\in\mathbb{R}} \frac{g_x(\alpha)}{g_x(\theta_{n,x})} = O_{\mathbb{P}}(1). \tag{3.14}$$

Let $\delta \in (0, 1/2)$ be small. The region over which the second supremum is taken will be split up into three regions, namely $F(x) < \delta$, $\delta \le F(x) \le 1-\delta$ and $F(x) > 1-\delta$, respectively. For the middle region we have for small enough η, because of Corollary 3.1, that almost surely for large n, $g_x(\theta_{n,x})$ is bounded away from 0; see proof of Lemma 3.3. Since $g_x(\alpha) \le \sup_{y\in\mathbb{R}} f(y) = f(x_1)$, we hence have

$$\sup_{\alpha<\eta}\sup_{\delta\le F(x)\le 1-\delta} \frac{g_x(\alpha)}{g_x(\theta_{n,x})} = O_{\mathbb{P}}(1).$$

In order to complete the proof of (3.14) we need to consider the regions $F(x) < \delta$ and $F(x) > 1-\delta$. Because of symmetry we will restrict ourselves to the region $F(x) < \delta$. Note that from the proof of Lemma 3.3 it follows that g_x is nondecreasing for $\alpha \le |F(x_1) - F(x)|$ and nonincreasing for $\alpha \ge |F(x_1) - F(x)|$. Therefore for $\alpha \ge V_{n,x}(\alpha)$ (when $\alpha \le V_{n,x}(\alpha)$ we use 1 as an upper bound), almost surely for large n

$$\begin{aligned}\sup_{\alpha<\eta}\sup_{F(x)<\delta} \frac{g_x(\alpha)}{g_x(\theta_{n,x})} &\le \sup_{\alpha<\eta}\sup_{F(x)<\delta} \frac{g_x(\alpha)}{g_x(V_{n,x}(\alpha))} \\ &\le \sup_{\alpha<\eta}\sup_{F(x)\le\alpha} \frac{g_x(\alpha)}{g_x(V_{n,x}(\alpha))} + \sup_{\alpha<\eta}\sup_{\alpha<F(x)<\delta} \frac{g_x(\alpha)}{g_x(V_{n,x}(\alpha))}\,. \end{aligned} \tag{3.15}$$

For x and α both small enough we have $g_x(\alpha) = f\left(F^{-1}\left(F(x)+\alpha\right)\right)$. Hence the second term in the right-hand side of (3.15) is equal to

$$\sup_{\alpha<\eta}\sup_{\alpha<F(x)<\delta} \frac{f\left(F^{-1}\left(F(x)+\alpha\right)\right)}{f\left(F^{-1}\left(F(x)+V_{n,x}(\alpha)\right)\right)} \le \sup_{F(x)<\delta} \frac{f\left(F^{-1}\left(2F(x)\right)\right)}{f\left(F^{-1}\left(F(x)\right)\right)}\,, \tag{3.16}$$

because $f \circ F^{-1}$ is increasing on $(0, F(x_1))$. It is immediate from Fact 3.1 that the right-hand side of (3.16) is bounded.

Similarly by Fact 3.1, the first term in the right-hand side of (3.15) is bounded from above by

$$\begin{aligned} &\sup_{\alpha<\eta} \sup_{F(x)\le\alpha} \frac{f\left(F^{-1}\left(2\alpha\right)\right)}{f\left(F^{-1}\left(F(x)+V_{n,x}(\alpha)\right)\right)} \\ &\quad \le \sup_{\alpha<\eta} \sup_{F(x)\le\alpha} \left(\frac{2\alpha}{F(x)+V_{n,x}(\alpha)} \cdot \frac{1-F(x)-V_{n,x}(\alpha)}{1-2\alpha} \right)^M , \end{aligned} \tag{3.17}$$

Because the second factor in the right-hand side of (3.17) is clearly bounded in probability, we need to show that

$$\sup_{\alpha<\eta} \sup_{F(x)\le\alpha} \frac{\alpha}{F(x)+V_{n,x}(\alpha)} \;=\; O_{\mathbb{P}}(1).$$

The proof of this is based on the following crucial inequality: with probability 1,

$$V_{n,x}(\alpha) \ge F(X_{(\lceil n\alpha \rceil)}) - F(x) \quad \text{for all } x \in \mathbb{R} \text{ and } 0 < \alpha < 1. \tag{3.18}$$

When proving this inequality, we assume $F(X_{(\lceil n\alpha \rceil)}) > F(x)$, otherwise there is nothing to prove. Using Lemma 3.6 and the Monotonicity property (Section 3.2), we see that we need to show

$$S_{n,x}(\alpha) \ge S_x(F(X_{(\lceil n\alpha \rceil)}) - F(x)).$$

From the definitions of S_x and $S_{n,x}$ we obtain

$$\begin{aligned} &S_x(F(X_{(\lceil n\alpha \rceil)}) - F(x)) \\ &\quad = \inf\{b-a : F(b)-F(a) \ge F(X_{(\lceil n\alpha \rceil)}) - F(x), x \in [a,b]\} \\ &\quad \le X_{(\lceil n\alpha \rceil)} - x \le S_{n,x}(\alpha), \end{aligned}$$

and hence (3.18).

Now (3.18) yields, almost surely,

$$\sup_{\alpha<\eta}\sup_{F(x)\le\alpha}\frac{\alpha}{F(x)+V_{n,x}(\alpha)}\le\sup_{\alpha<\eta}\frac{\alpha}{F(X_{(\lceil n\alpha\rceil)})}\le\sup_{0<\alpha<1}\frac{\alpha}{F(X_{(\lceil n\alpha\rceil)})}\ . \quad (3.19)$$

The denominator on the right is equal in distribution to the empirical quantile function of a sample of n independent uniform-$(0,1)$ variables. Hence it is well-known that the expression on the right in (3.19) is bounded in probability, see, e.g., Shorack and Wellner (1986), p. 419. This proves (3.14).

The proof of Theorem 3.2 is complete if we show that

$$\sup_{\alpha>1-\eta}\sup_{x\in\mathbb{R}}\frac{g_x(\alpha)}{g_x(\theta_{n,x})}=O_{\mathbb{P}}(1). \quad (3.20)$$

For symmetry reasons we can restrict x to $(-\infty,F^{-1}(1/2)]$. For large enough α, λ_α is defined and we can write g_x as follows:

$$g_x(\alpha)=\begin{cases} f(F^{-1}(F(x)+\alpha)) & \text{for } x<F^{-1}((1-\alpha)\lambda_\alpha)\\ f(F^{-1}((1-\alpha)\lambda_\alpha)) & \text{for } F^{-1}((1-\alpha)\lambda_\alpha)\le x\le F^{-1}(1/2)\ .\end{cases}$$

For small enough η and $\alpha\le V_{n,x}(\alpha)$ (again, when $\alpha\ge V_{n,x}(\alpha)$ we can use 1 as an upper bound) we obtain

$$\begin{aligned}\sup_{\alpha>1-\eta}\sup_{F(x)\le\frac12}\frac{g_x(\alpha)}{g_x(\theta_{n,x})} \le{}& \sup_{\alpha>1-\eta}\sup_{F(x)<(1-V_{n,x}(\alpha))\lambda_{V_{n,x}(\alpha)}}\frac{g_x(\alpha)}{g_x(V_{n,x}(\alpha))} \quad (3.21)\\ &\vee\sup_{\alpha>1-\eta}\sup_{(1-V_{n,x}(\alpha))\lambda_{V_{n,x}(\alpha)}\le F(x)<(1-\alpha)\lambda_\alpha}\frac{g_x(\alpha)}{g_x(V_{n,x}(\alpha))}\\ &\vee\sup_{\alpha>1-\eta}\sup_{(1-\alpha)\lambda_\alpha\le F(x)\le\frac12}\frac{g_x(\alpha)}{g_x(V_{n,x}(\alpha))}\ .\end{aligned}$$

The last term in the right-hand side of (3.21) can be bounded from above by applying Fact 3.1:

$$\sup_{\alpha>1-\eta}\sup_{(1-\alpha)\lambda_\alpha\le F(x)\le\frac12}\frac{g_x(\alpha)}{g_x(V_{n,x}(\alpha))}$$

$$
\begin{aligned}
&= \sup_{\alpha>1-\eta} \sup_{(1-\alpha)\lambda_\alpha \le F(x) \le \frac{1}{2}} \frac{f\left(F^{-1}\left((1-\alpha)\lambda_\alpha\right)\right)}{f\left(F^{-1}\left(\left(1-V_{n,x}(\alpha)\right)\lambda_{V_{n,x}(\alpha)}\right)\right)} \\
&\le \sup_{\alpha>1-\eta} \sup_{x\in\mathbb{R}} \left(\frac{(1-\alpha)\lambda_\alpha}{\left(1-V_{n,x}(\alpha)\right)\lambda_{V_{n,x}(\alpha)}} \, \frac{1-\left(1-V_{n,x}(\alpha)\right)\lambda_{V_{n,x}(\alpha)}}{1-(1-\alpha)\lambda_\alpha} \right)^M ,
\end{aligned}
$$

which is easily seen to be $O_{\mathbb{P}}(1)$, using Lemma 3.7 and condition (C).

The first term in the right-hand side of (3.21) is equal to

$$
\sup_{\alpha>1-\eta} \sup_{F(x)<(1-V_{n,x}(\alpha))\lambda_{V_{n,x}(\alpha)}} \frac{f\left(F^{-1}\left(F(x)+\alpha\right)\right)}{f\left(F^{-1}\left(F(x)+V_{n,x}(\alpha)\right)\right)},
$$

which is, since f is decreasing for large values and because of Fact 3.1, bounded from above by

$$
\begin{aligned}
&\sup_{\alpha>1-\eta} \sup_{F(x)<(1-V_{n,x}(\alpha))\lambda_{V_{n,x}(\alpha)}} \frac{f\left(F^{-1}\left(F(x)+\alpha\right)\right)}{f\left(F^{-1}\left(\left(1-V_{n,x}(\alpha)\right)\lambda_{V_{n,x}(\alpha)}+V_{n,x}(\alpha)\right)\right)} \\
&\quad \le \sup_{\alpha>1-\eta} \sup_{x\in\mathbb{R}} \left(\frac{\left(1-V_{n,x}(\alpha)\right)\lambda_{V_{n,x}(\alpha)}+V_{n,x}(\alpha)}{\alpha} \right. \\
&\qquad\qquad \left. \frac{1-\alpha}{\left(1-V_{n,x}(\alpha)\right)\left(1-\lambda_{V_{n,x}(\alpha)}\right)} \right)^M . \qquad (3.22)
\end{aligned}
$$

Again, Lemma 3.7 and condition (C) yield that this term is $O_{\mathbb{P}}(1)$.

The middle term in the right-hand side of (3.21), rewritten as (using $f(F^{-1}((1-\beta)\lambda_\beta)) = f(F^{-1}((1-\beta)\lambda_\beta+\beta))$)

$$
\sup_{\alpha>1-\eta} \sup_{(1-V_{n,x}(\alpha))\lambda_{V_{n,x}(\alpha)} \le F(x) < (1-\alpha)\lambda_\alpha} \frac{f\left(F^{-1}\left(F(x)+\alpha\right)\right)}{f\left(F^{-1}\left(\left(1-V_{n,x}(\alpha)\right)\lambda_{V_{n,x}(\alpha)}+V_{n,x}(\alpha)\right)\right)} ,
$$

is bounded by the right-hand side of (3.22). This completes the proof of (3.20). □

Chapter 4

The Half-Half plot

[Based on joint work with J.H.J. Einmahl, *The Half-Half plot*, CentER discussion paper 2009-77.]

Abstract: The Half-Half (HH) plot is a new graphical method to investigate qualitatively the shape of a regression curve. The empirical HH-plot counts observations in the lower and upper quarter of a strip that moves horizontally over the scatter plot. The plot displays jumps clearly and reveals further features of the regression curve. We prove a functional central limit theorem for the empirical HH-plot, with rate of convergence $1/\sqrt{n}$. In a simulation study the good performance of the plot is demonstrated. The method is also applied to a case study.

Keywords: Data analysis, functional central limit theorem, graphical methods, jump detection, nonparametric regression.

4.1 Introduction

Assume the pairs $(X, Y), (X_1, Y_1), \ldots, (X_n, Y_n)$, $n \in \mathbb{N}$, are independent and identically distributed (iid) with bivariate distribution function (df) F. From the regression perspective we can define $\varepsilon = Y - m(X)$ where m, the nonparametric regression function, is a location functional (like the median, mean or

mode) applied to the conditional distribution of $Y|X = x$. Equivalently, this can be written in the 'standard form':

$$Y = m(X) + \varepsilon.$$

A first step in data analysis is exploratory diagnostics. Using a good graphical representation of the data sheds light on their main features. The standard procedure for depicting m is kernel estimation, which produces a smooth regression function. When there are sudden changes in m, such as jumps, those are forced into the smoothed picture and therefore ignored. More refined procedures that allow for and detect jumps of m have been introduced and will be discussed later.

In this paper, we present a novel, nonparametric, computationally fast method for explaining regression curves. It displays important features of a regression curve, such as jumps and in- or decreases. If it is a goal to search for jumps, then an ad hoc estimator (introduced in Section 4.3.2) can be used to find their locations. These procedures impose no particular model on the data; also the regression function m need not be estimated.

Let F be absolutely continuous with density f. Denote the corresponding probability measure with P. Write F_1, F_2 for the marginals of F and Q_1, Q_2 for their (left-continuous) inverse or quantile functions. The empirical counterparts of these functions are denoted with a subscript n, in particular F_n denotes the empirical df of the $(X_i, Y_i), i = 1, \ldots, n$, i.e.

$$F_n(x, y) = \tfrac{1}{n} \sum_{i=1}^{n} 1_{(-\infty,x]\times(-\infty,y]}(X_i, Y_i), \quad -\infty < x, y \leq \infty.$$

Fix $\alpha \in (0, \frac{1}{2})$. For $x \in (Q_1(\alpha), Q_1(1-\alpha))$ define the vertical α-strip centered at x by

$$S_\alpha(x) := \{(u, v) \in \mathbb{R}^2 : Q_1(F_1(x) - \alpha) \leq u \leq Q_1(F_1(x) + \alpha)\}.$$

Consider the univariate distribution function on this strip

$$G_{x,\alpha}(y) := \tfrac{1}{2\alpha}\left[F(Q_1(F_1(x)+\alpha),y) - F(Q_1(F_1(x)-\alpha),y)\right], \quad y \in \mathbb{R}.$$

The corresponding quantile function is denoted by $Q_{x,\alpha}$. For convenience, we write in the sequel $x^- = Q_1(F_1(x)-\alpha)$ and $x^+ = Q_1(F_1(x)+\alpha)$.

Definition 4.1. Let the coverage $\alpha \in (0, \frac{1}{2})$ be fixed. The Half-Half (HH) value for $x \in (Q_1(\alpha), Q_1(1-\alpha))$ is defined by

$$\begin{aligned} H_\alpha(x) &= F\left(x, Q_{x,\alpha}\left(\tfrac{1}{4}\right)\right) - F\left(x^-, Q_{x,\alpha}\left(\tfrac{1}{4}\right)\right) \\ &\quad - \left(F\left(x^+, Q_{x,\alpha}\left(\tfrac{3}{4}\right)\right) - F\left(x, Q_{x,\alpha}\left(\tfrac{3}{4}\right)\right)\right) + \tfrac{1}{2}\alpha. \end{aligned}$$

The HH-value is obtained by first vertically dividing the strip $S_\alpha(x)$ into two halves of equal mass: the middle half and the outer half (lower and upper quarter). Then the mass of the lower quarter (i.e. below $Q_{x,\alpha}(\frac{1}{4})$) that is in the left half of the strip (i.e. in $(x^-, x] \times \mathbb{R}$) is added to the mass of the upper quarter that is in the right half of the strip. To standardize the statistic, $\frac{1}{2}\alpha$ is subtracted; this is the sum of these masses corresponding to X and Y being independent. Hence, in that case $H_\alpha \equiv 0$. In general, $H_\alpha \in [-\frac{1}{2}\alpha, \frac{1}{2}\alpha]$. Note that

$$\begin{aligned} H_\alpha(x) &= F\left(x, Q_{x,\alpha}\left(\tfrac{1}{4}\right)\right) - F\left(x^-, Q_{x,\alpha}\left(\tfrac{1}{4}\right)\right) \\ &\quad + F\left(x, Q_{x,\alpha}\left(\tfrac{3}{4}\right)\right) - F\left(x^-, Q_{x,\alpha}\left(\tfrac{3}{4}\right)\right) - \alpha. \end{aligned}$$

Definition 4.2. Let $\alpha \in (0, \frac{1}{2})$. The Half-Half (HH) plot is the graph of the function

$$x \mapsto H_\alpha(x),\ x \in (Q_1(\alpha), Q_1(1-\alpha)).$$

Define for $x \in (Q_{n,1}(\alpha), Q_{n,1}(1-\alpha))$ the empirical counterpart of $G_{x,\alpha}$ by

$$G_{n,x,\alpha}(y) = \tfrac{1}{2\alpha}\left[F_n(Q_{n,1}(F_{n,1}(x)+\alpha),y) - F_n(Q_{n,1}(F_{n,1}(x) - \tfrac{\lceil n\alpha \rceil}{n}),y)\right], \quad y \in \mathbb{R},$$

where $\lceil \cdot \rceil$ denotes the ceiling function. Write $X_{[1]} \leq X_{[2]} \leq \cdots \leq X_{[n]}$ for the order statistics of the $X_i, i = 1, \ldots, n$. It is easily seen that

$$G_{n,x,\alpha}(y) = \tfrac{1}{2\alpha} \left[F_n(X_{[i+\lceil n\alpha \rceil]}, y) - F_n(X_{[i-\lceil n\alpha \rceil]}, y) \right],$$

where i is such that $X_{[i]} \leq x < X_{[i+1]}$. Denote the quantile function of $G_{n,x,\alpha}$ with $Q_{n,x,\alpha}$. The HH-statistic is for $x \in (Q_{n,1}(\alpha), Q_{n,1}(1-\alpha))$ defined to be the empirical version of the HH-value:

$$\begin{aligned} H_{n,\alpha}(x) \;=\; & F_n\left(x, Q_{n,x,\alpha}\left(\tfrac{1}{4}\right)\right) - F_n\left(Q_{n,1}\left(F_{n,1}(x) - \tfrac{\lceil n\alpha \rceil}{n}\right), Q_{n,x,\alpha}\left(\tfrac{1}{4}\right)\right) \\ & + F_n\left(x, Q_{n,x,\alpha}\left(\tfrac{3}{4}\right)\right) - F_n\left(Q_{n,1}\left(F_{n,1}(x) - \tfrac{\lceil n\alpha \rceil}{n}\right), Q_{n,x,\alpha}\left(\tfrac{3}{4}\right)\right) \;-\; \alpha \\ = \; & F_n\left(X_{[i]}, Q_{n,x,\alpha}\left(\tfrac{1}{4}\right)\right) - F_n\left(X_{[i-\lceil n\alpha \rceil]}, Q_{n,x,\alpha}\left(\tfrac{1}{4}\right)\right) \\ & + F_n\left(X_{[i]}, Q_{n,x,\alpha}\left(\tfrac{3}{4}\right)\right) - F_n\left(X_{[i-\lceil n\alpha \rceil]}, Q_{n,x,\alpha}\left(\tfrac{3}{4}\right)\right) \;-\; \alpha. \end{aligned}$$

The empirical HH-plot is defined to be the graph of the function

$$x \mapsto H_{n,\alpha}(x), \quad x \in (Q_{n,1}(\alpha), Q_{n,1}(1-\alpha)).$$

Jumps of m are depicted in the HH-plot by a high positive (jump up) or negative (jump down) value. In the first example, we consider $m_1(x) = \mathbb{1}_{[0.5,\infty)}(x)$, with X uniformly distributed on $(0,1)$ and ε standard normal and independent of X, see Figure 4.1. The jump is clearly indicated by the large HH-value and -statistic at $x = 0.5$. Additionally, the HH-plot reveals that the

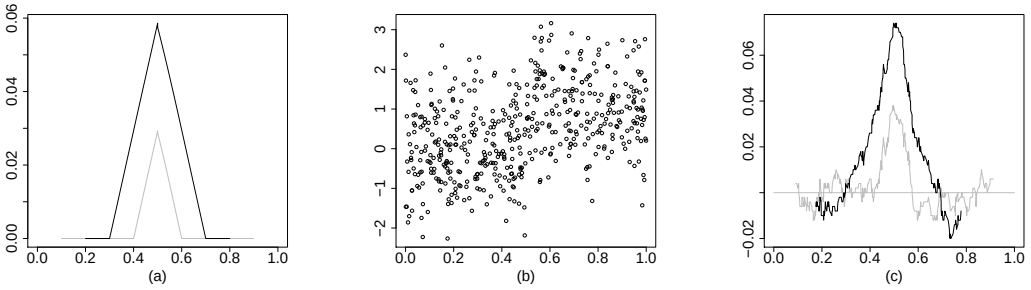

Figure 4.1: Theoretical (a) and empirical (c) HH-plot with $\alpha = 0.1$ (grey line) and $\alpha = 0.2$ (black line) of a sample of size $n = 500$ (b) of m_1 with ε standard normal and $X \sim UN(0,1)$.

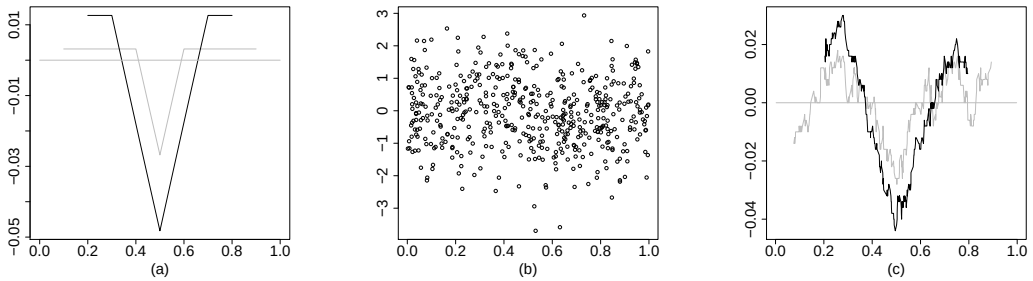

Figure 4.2: Theoretical (a) and empirical (c) HH-plot with $\alpha = 0.1$ (grey line) and $\alpha = 0.2$ (black line) of a sample of size $n = 500$ (b) of m_2 with ε standard normal and $X \sim UN(0, 1)$.

regression curve is constant before and after the jump, which is indicated by values of $H_\alpha(x)$ and $H_{n,\alpha}(x)$ close to 0.

Another important feature of the HH-plot can be observed in Figure 4.2. Let X and ε be as above, but $m_2(x) = x - \mathbb{1}_{[0.5,\infty)}(x)$. The jump down is indicated by large negative values of $H_\alpha(0.5)$ and $H_{n,\alpha}(0.5)$, whereas the positive values indicate that m_2 is increasing. Hence the HH-plot can be used for detecting jumps as well as continuous increases or decreases of the regression curve.

It is advised to calculate the HH-plot for two values of α. Changes in the regression are typically easier detected by using larger α-strips. But because the empirical HH-statistic is only defined for $x \in (Q_{n,1}(\alpha), Q_{n,1}(1-\alpha))$, a larger α reduces the range. A natural maximum here seems to be $\alpha = 0.25$. Hence, in order to get an impression of how the regression curve behaves in the beginning and the end, additionally a HH-plot with a small α should be depicted. A rule of thumb is that for $n \geq 200$ the larger α should be taken 0.2, and the smaller α can be chosen such that $n\alpha \geq 25$; for $n < 200$ we advise to draw the HH-plot for only one α, for instance 0.2.

There exists a large body of literature on the specific topic of estimating jump points in nonparametric regression. A good overview can, for example, be found in Gijbels et al. (2007). The most common approach is to compare left- and right-sided estimators of the regression function at point

x. These estimators are attained by kernel methods [amongst others, Müller (1992), Hall and Titterington (1992), and Wu and Chu (1993)] or local polynomial regression [amongst others, McDonald and Owen (1986), Loader (1996), Horváth and Kokoszka (2002), Grégoire and Hamrouni (2002), and Gijbels et al. (2007)]. Some papers propose a two-step procedure, where in a first step the location of jump is estimated and in a second step the curve is fitted [cf. Gijbels et al. (1999), Sánchez-Borrego et al. (2006)]. Wavelet methods are, amongst others, employed in Wang (1995), Antonidiadis and Gijbels (2002) and Park and Kim (2006). In Kim and Marron (2006), another graphical method, the SiZer for jump detection, is introduced. The SiZer is a kernel-based approach which combines multiple bandwidths in one plot. Its main goal is to get a first, general overview of the regression curve.

In the papers mentioned above, the specific assumptions on the regression curve and the errors play an important role. In contrast, we have no direct assumptions on m or ε. In Dempfle and Stute (2002) an approach based on U-statistics is used. As in the present paper, this approach requires minimal assumptions on the model and avoids smoothing for estimating the location of the jump.

The paper is organized as follows. The asymptotic behavior of the HH-statistic is expounded in Section 4.2, where it is shown that the rate of convergence of the empirical HH-statistic to the theoretical one is $1/\sqrt{n}$, uniformly in x. The limiting process of the HH-value is also used to facilitate the depiction of jumps. In Section 4.3 three simulation studies are presented and an ad hoc estimator for a jump location is introduced in Section 4.3.2. A real data application can be found in Section 4.4. The paper is completed by a section containing the proof of the result of Section 4.2.

4.2 Asymptotic results

Define for $n \in \mathbb{N}$ the empirical process indexed by points as

$$U_n(x,y) := n^{\frac{1}{2}}\big\{F_n(x,y) - F(x,y)\big\}, \quad -\infty < x, y \leq \infty.$$

Furthermore, let B_F be a bounded, mean zero Gaussian process on $(-\infty,\infty]^2$ that is uniformly continuous and has covariance function $F(x_1 \wedge x_2, y_1 \wedge y_2) - F(x_1,y_1)F(x_2,y_2)$, $(x_1,y_1),(x_2,y_2) \in (-\infty,\infty]^2$. Then, by the functional central limit theorem for U_n and the Skorohod representation theorem, there exist $\tilde{B}_F \stackrel{d}{=} B_F$ and a sequence $\tilde{U}_n \stackrel{d}{=} U_n$, $n \in \mathbb{N}$, such that

$$\sup_{-\infty<x,y\leq\infty} \left|\tilde{U}_n(x,y) - \tilde{B}_F(x,y)\right| \to 0 \quad \text{a.s. as } n \to \infty. \tag{4.1}$$

Henceforth we will use (4.1), but drop the tildes from the notation.

We will use the following assumptions:

(A) Let $\mathcal{S} := \{(x,y) \in \mathbb{R}^2 : f(x,y) > 0\} = (x_*, x^*) \times \mathbb{R}$ for some $-\infty \leq x_* < x^* \leq \infty$ and F be of class C^2 on $\mathcal{S}$ and f_2 be bounded.

Hence, the function $G_{x,\alpha}(\cdot)$ is increasing and continuous on $(Q_1(\alpha), Q_1(1-\alpha))$; also we can write $F'_x(x,y) := \frac{\partial}{\partial x}F(x,y) = \int_{-\infty}^{y} f(x,v)dv$ and $F'_y(x,y) := \frac{\partial}{\partial y}F(x,y) = \int_{-\infty}^{x} f(u,y)du$.

For the second assumption we first introduce some more notation: Let $\tilde{F}(u,y) := F(Q_1(u),y)$ denote the in the first coordinate uniformized distribution function, and write $\tilde{F}'_x(u,y) := \frac{\partial}{\partial u}\tilde{F}(u,y) = \frac{F'_x(Q_1(u),y)}{f_1(Q_1(u))}$ and $\tilde{F}'_y(u,y) := \frac{\partial}{\partial y}\tilde{F}(u,y) = F'_y(Q_1(u),y)$.

(B) $\tilde{F}'_x$ and $\tilde{F}'_y$ are uniformly continuous on $(0,1) \times \mathbb{R}$.

Let $\alpha \in (0, \frac{1}{2})$ and define $I_n := (Q_1(\alpha), Q_1(1-\alpha)) \cap (Q_{n,1}(\alpha), Q_{n,1}(1-\alpha))$ and $I_0 := (Q_1(\alpha), Q_1(1-\alpha))$. Denote the HH-process with

$$V_{n,\alpha}(x) := n^{\frac{1}{2}}\big(H_{n,\alpha}(x) - H_\alpha(x)\big), \quad x \in I_n,$$

and for $s \in \{\frac{1}{4}, \frac{3}{4}\}$ define

$$\begin{aligned} B_{\alpha,s}(x) \;=\; & B_F\big(x, Q_{x,\alpha}(s)\big) \;-\; B_F\big(x^-, Q_{x,\alpha}(s)\big) \\ & - \frac{F'_x\big(x^-, Q_{x,\alpha}(s)\big)}{f_1\big(x^-\big)}\big[B_F\big(x,\infty\big) \;-\; B_F\big(x^-,\infty\big)\big] \\ & + \frac{F'_y\big(x^-, Q_{x,\alpha}(s)\big) \;-\; F'_y\big(x, Q_{x,\alpha}(s)\big)}{F'_y\big(x^+, Q_{x,\alpha}(s)\big) - F'_y\big(x^-, Q_{x,\alpha}(s)\big)}\Big\{B_F\big(x^+, Q_{x,\alpha}(s)\big) - B_F\big(x^-, Q_{x,\alpha}(s)\big) \\ & \quad + \frac{F'_x\big(x^+, Q_{x,\alpha}(s)\big)}{f_1\big(x^+\big)}\;\big[B_F(x,\infty) \;-\; B_F\big(x^+,\infty\big)\big] \\ & \quad - \frac{F'_x\big(x^-, Q_{x,\alpha}(s)\big)}{f_1\big(x^-\big)}\;\big[B_F(x,\infty) \;-\; B_F\big(x^-,\infty\big)\big]\Big\}, \quad x \in I_0. \end{aligned} \tag{4.2}$$

Theorem 4.1. *Under the assumptions (A) and (B), we have for all* $\alpha \in (0, \frac{1}{2})$, *on the probability space of (4.1),*

$$\sup_{x \in I_n} \big|V_{n,\alpha}(x) - \big(B_{\alpha,1/4}(x) + B_{\alpha,3/4}(x)\big)\big| \to 0 \quad \textit{a.s. as } n \to \infty.$$

Remark 4.1. For X and Y independent (m is constant), the distribution of the limiting process from Theorem 4.1 for fixed $x \in I_0$ simplifies to a $N(0, \frac{1}{4}\alpha)$-distribution.

Proof of Remark 4.1 In this case the limiting process specializes to

$$\begin{aligned} B_{\alpha,1/4}(x) + B_{\alpha,3/4}(x) \;=\; & \tfrac{1}{2}\big\{\big[B_F(x, Q_2\left(\tfrac{1}{4}\right)) - B_F(x^-, Q_2\left(\tfrac{1}{4}\right))\big] \\ & - \big[B_F(x^+, Q_2\left(\tfrac{1}{4}\right)) - B_F(x, Q_2\left(\tfrac{1}{4}\right))\big] \;+\; \big[B_F(x, Q_2\left(\tfrac{3}{4}\right)) - B_F(x^-, Q_2\left(\tfrac{3}{4}\right))\big] \\ & - \big[B_F(x^+, Q_2\left(\tfrac{3}{4}\right)) - B_F(x, Q_2\left(\tfrac{3}{4}\right))\big] \;-\; \big[B_F(x,\infty) - B_F(x^-,\infty)\big] \\ & + \big[B_F(x^+,\infty) - B_F(x,\infty)\big]\big\}. \end{aligned}$$

Taking the sets $A_s := (x^-, x] \times (-\infty, Q_2(s)]$, $B_s := (x, x^+] \times (-\infty, Q_2(s)]$, $C_s := (x^-, x] \times (Q_2(s), \infty)$ and $D_s := (x, x^+] \times (Q_2(s), \infty)$, as depicted in Figure 4.3, as well as $E^- := (x^-, x] \times \mathbb{R}$ and $E^+ := (x, x^+] \times \mathbb{R}$, into account, we have $P(A_s) = P(B_s) = s\alpha$, $P(C_s) = P(D_s) = (1-s)\alpha$, $P(E^-) = P(E^+) = \alpha$.

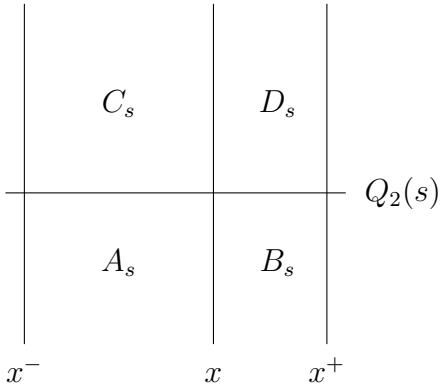

Figure 4.3: Regions taken into account for the calculation of the limiting process at x, where $s \in \{\frac{1}{4}, \frac{3}{4}\}$.

Extending the definition of B_F to semi-infinite rectangles in the usual way, we get

$$\begin{aligned} & B_{\alpha,1/4}(x) + B_{\alpha,3/4}(x) \qquad (4.3) \\ = \; & \tfrac{1}{2}\left\{B_F(A_{1/4}) - B_F(B_{1/4}) + B_F(A_{3/4}) - B_F(B_{3/4}) - B_F(E^-) + B_F(E^+)\right\} \\ = \; & \tfrac{1}{2}\left\{B_F(A_{1/4}) - B_F(B_{1/4}) - B_F(C_{3/4}) + B_F(D_{3/4})\right\}. \end{aligned}$$

Let W_F be the Wiener process defined on semi-infinite rectangles R by $W_F(R) = B_F(R) + P(R)Z$, where $Z \sim N(0,1)$ and B_F are independent. Then the right-hand side of (4.3) can be rewritten as

$$\tfrac{1}{2}\left\{W_F(A_{1/4}) - W_F(B_{1/4}) - W_F(C_{3/4}) + W_F(D_{3/4})\right\}, \qquad (4.4)$$

which is easily seen to be $N(0, \frac{1}{4}\alpha)$-distributed. □

4.3 Simulation study

For an easier interpretation of the HH-plot, we add a horizontal band to the picture. If the empirical HH-plot escapes the band, this indicates jumps or steep in- or decreases of the regression curve. In addition, this band gives a standard to assess the relative magnitude of the HH-statistic. The band is

obtained by calculating high quantiles of

$$\sup_{x \in I_0} \left| B_{\alpha,1/4}(x) + B_{\alpha,3/4}(x) \right|, \tag{4.5}$$

in case X and Y are independent; (4.4) is useful for this calculation. These quantiles are denoted q_α (see Table 4.1) and have to be divided by $\sqrt{n}$ and $-\sqrt{n}$, respectively, in order to obtain the upper and lower boundary of the band.

α	0.05	0.1	0.15	0.2	0.25
$q_\alpha(0.9)$	0.40	0.53	0.61	0.67	0.72
$q_\alpha(0.95)$	0.42	0.56	0.66	0.73	0.79
$q_\alpha(0.99)$	0.47	0.64	0.75	0.84	0.91

Table 4.1: The 0.90, 0.95 and 0.99 quantiles of the random variable in (4.5), for $\alpha \in \{0.05, 0.1, 0.15, 0.2, 0.25\}$.

4.3.1 X and Y independent

First, we consider the case of X,Y being independent, with sample sizes of $n = 250$ and 500 for coverage levels $\alpha = 0.1$ and 0.2. For each n, 10,000 samples are taken. Table 4.2 provides the fraction of exceedances of $q_\alpha(0.95)/\sqrt{n}$ by $\max\{|H_{n,\alpha}(x)| : Q_{n,1}(\alpha) < x < Q_{n,1}(1-\alpha)\}$. We see that these numbers are close to, but somewhat lower than, $1 - 0.95 = 0.05$. This might be due to the fact that for fixed x, the effective sample size (i.e. the number of observations in the lower and upper quarter of the strip) is $n\alpha$, which can be as small as 25. Indeed, for $n = 500$ and $\alpha = 0.2$, we see that the asymptotic quantiles are quite accurate.

4.3.2 Regression functions m_1 and m_2 of Section 4.1

Next, we consider m_1 and m_2 as in Section 4.1. We simulate from the models presented there, and additionally take the parameters $\sigma = 0.1$ and $\sigma = 0.5$ for

	$\max(\lvert H_{n,\alpha}(x)\rvert) \geq q_\alpha(0.95)/\sqrt{n}$	
n	$\alpha = 0.1$	$\alpha = 0.2$
250	0.025	0.032
500	0.037	0.044

Table 4.2: Simulated Type I error probabilities for X and Y independent.

the normal distribution of the errors into account. Sample sizes and numbers of replications are taken as above. As before, the upwards (m_1) and downwards (m_2) fraction of exceedances of $q_\alpha/\sqrt{n}$ and $-q_\alpha/\sqrt{n}$, respectively, is measured. In addition, the fraction of how often there is indeed a jump up of m_1 detected at $x = 0.5$, thus how often the empirical HH-plot at $x = 0.5$ exceeds $q_\alpha(0.95)/\sqrt{n}$, is reported. Similarly, this fraction is also given for the jump down of m_2.

m_1		$\max(H_{n,\alpha}(x)) \geq q_\alpha(0.95)/\sqrt{n}$		$H_{n,\alpha}(0.5) \geq q_\alpha(0.95)/\sqrt{n}$		$\mathrm{sd}(\hat{\theta})$	
σ	n	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.2$
0.1	250	1.000	1.000	1.000	1.000	0.007	0.007
	500	1.000	1.000	1.000	1.000	0.003	0.003
0.5	250	0.993	1.000	0.979	1.000	0.011	0.011
	500	1.000	1.000	1.000	1.000	0.006	0.006
1	250	0.412	0.957	0.243	0.858	0.069	0.038
	500	0.902	1.000	0.766	0.998	0.027	0.020

Table 4.3: Simulation study as described in Section 4.3.2 for the regression function m_1.

The exceedance rates close to 100% in Tables 4.3 and 4.4 show that the HH-plot is a very good tool to detect jumps. Only at $\sigma = 1$, when the picture is very blurry, the effective sample size of $n\alpha = 25$ ($n = 250$, $\alpha = 0.1$) is too small and the HH-plot has less power.

Last, an ad hoc estimator $\hat{\theta}$ of the jump location is given, by taking in every simulation the mean of $\mathrm{argmax}\{H_{n,\alpha}(x) : x \in (Q_{n,1}(\alpha), Q_{n,1}(1-\alpha))\}$ (and, for m_2, similarly the mean of the argmin). Due to symmetry, this estimator is unbiased. The standard deviations, presented in the last two columns of

m_2		$\min(H_{n,\alpha}(x)) \leq -q_\alpha(0.95)/\sqrt{n}$		$H_{n,\alpha}(0.5) \leq -q_\alpha(0.95)/\sqrt{n}$		$\mathrm{sd}(\hat{\theta})$	
σ	n	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.2$
0.1	250	1.000	1.000	1.000	1.000	0.005	0.004
	500	1.000	1.000	1.000	1.000	0.002	0.002
0.5	250	0.965	1.000	0.931	1.000	0.012	0.012
	500	1.000	1.000	1.000	1.000	0.006	0.006
1	250	0.289	0.737	0.159	0.569	0.070	0.038
	500	0.777	0.985	0.616	0.958	0.026	0.020

Table 4.4: Simulation study as described in Section 4.3.2 for the regression function m_2.

Tables 4.3 and 4.4, show that this approach of averaging the locations of the extrema leads to a good estimator of the jump location.

4.3.3 Regression function as in Gijbels et al. (1999)

The final simulation is an example from Gijbels et al. (1999), with regression function

$$m_3(x) = \begin{cases} \exp\{-2(x-0.35)\} - 1 & \text{if} \quad x \in [0, 0.35) \\ \exp\{-2(x-0.35)\} & \text{if} \quad x \in [0.35, 0.65) \\ \exp\{2(x-0.65)\} + \exp\{-0.6\} - 2 & \text{if} \quad x \in [0.65, 1]. \end{cases} \tag{4.6}$$

It is depicted in Figure 4.4(a); a scatter plot with $X \sim UN(0,1)$, ε standard normal and independent of X, and $n = 250$ is given in Figure 4.4(b).

In Figure 4.5, the theoretical HH-plots for m_3 with $X \sim UN(0,1)$ and $\varepsilon \sim N(0, \sigma^2)$, $\sigma \in \{0.1, 0.5, 1\}$, for both $\alpha = 0.1$ and $\alpha = 0.2$, are given. The features of the HH-plot are nicely shown in Figure 4.5(a). The possible range of $H_{0.1}$ is $[-0.05, 0.05]$, and that of $H_{0.2}$ is $[-0.10, 0.10]$, and in both cases it is fully utilized. The negative slope of the first part of m_3 is depicted in the negative, almost horizontal line of the HH-plot for $\alpha = 0.1$ and the negative values of $H_{0.2}$. The latter one starts at $x = 0.2$ and hence already takes the jump upwards into account ($2\alpha = 0.4 > 0.35$), which is depicted by

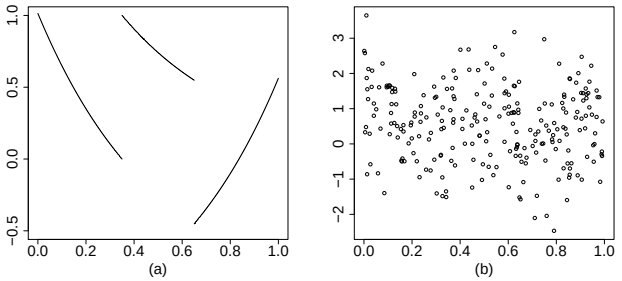

Figure 4.4: Panel (a) shows m_3 as given in (4.6) and panel (b) one simulation of m_3 with ε standard normal, $X \sim UN(0,1)$ and $n = 250$.

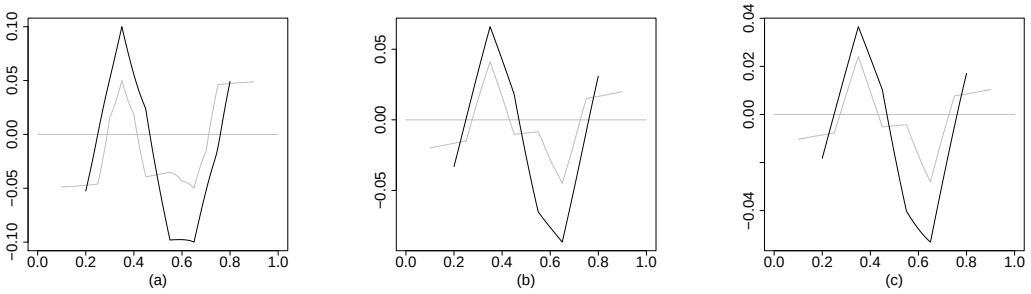

Figure 4.5: The theoretical HH-plot for m_3 is shown for various standard deviations ((a) $\sigma = 0.1$, (b) $\sigma = 0.5$ and (c) $\sigma = 1$) of the normal error, for $\alpha = 0.1$ (grey line) and $\alpha = 0.2$ (black line).

the increasing HH-values. For $\alpha = 0.1$, this increase starts at $x = 0.35 - \alpha = 0.25$. The jump point at $x = 0.35$ is for both coverage levels α depicted by a distinct local maximum. The regression curve after this jump is again decreasing, hence the HH-values become negative, even before $x^- > 0.35$. For $x \in (0.45, 0.55)$, the plot of $H_{0.1}$ only depicts the negative slope, and hence almost reaches its minimal possible value. The jump at $x = 0.65$ is then difficult to see from the HH-plot, because the HH-values for both $\alpha = 0.1$ and $\alpha = 0.2$ only change slightly in a region left of the jump point. The last part of m_3 has a positive slope, hence the HH-values become positive again.

Figures 4.5(b) and (c) bear similar features as (a). The jumps are indicated by the extrema of the HH-plot, the steep in- and decreases by positive/negative HH-values. But in contrast to panel (a), the jumps down of the regression

curve are depicted by distinct minima of the HH-plot. Also observe that the range of the HH-values is getting smaller when σ increases.

In Table 4.5, the outcomes of the simulations are presented, with columns as described for m_1 and m_2. These outcomes are very similar to the outcomes of the simulations for m_1 and m_2; in general the HH-plot depicts the main features of the regression curve very well, but again the combination of a smaller $n\alpha$ with a larger σ results in a lower power. Furthermore, the biases and standard deviations of the extrema of the HH-statistic are presented in Table 4.5. [Here we have restricted $\operatorname{argmax} H_{n,\alpha}$ in the definition of $\hat{\theta}_{max}$ to $(Q_{n,1}(\alpha), \frac{1}{2}]$ (and similarly $\operatorname{argmin} H_{n,\alpha}$ to $[\frac{1}{2}, Q_{n,1}(1-\alpha))$) in order not to confuse the steep increase for large x with the jump at $x = 0.35$.] The estimator of the jump up (at $x = 0.35$) is indicated as $\hat{\theta}_{max}$, that of the jump down (at $x = 0.65$) with $\hat{\theta}_{min}$. For comparison with Gijbels et al. (1999), we also include $n = 100$. As expected from the above discussion of Figure 4.5, $\hat{\theta}_{min}$ has a negative bias and a larger standard deviation. It is remarkable, but explainable through Figure 4.5, that $\hat{\theta}_{min}$ is performing better in the more 'difficult' setup of $\sigma = 0.5$ than in that of $\sigma = 0.1$.

Although estimation is not our main goal, our estimator compares well with Gijbels et al. (1999). We liken our best results to the best results of Gijbels et al. (1999). The estimators for the jump locations there will be denoted in the following with $\hat{\theta}^G_{max}$ for the first jump at $x = 0.35$, and $\hat{\theta}^G_{min}$ for the second jump at $x = 0.65$. The absolute values of the biases of our $\hat{\theta}_{max}$ are always smaller or equal than those of $\hat{\theta}^G_{max}$; the standard deviations are similar. For the jump down at $x = 0.65$, σ plays a role. For $\sigma = 0.1$, the absolute values of the biases and the standard deviations of $\hat{\theta}_{min}$ are both greater or equal than those of $\hat{\theta}^G_{min}$. But with σ increased, $\hat{\theta}_{min}$ has similar biases and even smaller standard deviations than $\hat{\theta}^G_{min}$. To conclude, our ad hoc estimator is a competing method to detect jump locations, especially recommended when the picture is blurry.

m_3		$\max(H_{n,\alpha}(x)) \geq$ $q_\alpha(0.95)/\sqrt{n}$		$H_{n,\alpha}(0.35) \geq$ $q_\alpha(0.95)/\sqrt{n}$		$\min(H_{n,\alpha}(x)) \leq$ $-q_\alpha(0.95)/\sqrt{n}$		$H_{n,\alpha}(0.65) \leq$ $-q_\alpha(0.95)/\sqrt{n}$	
σ	n	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.2$
0.1	250	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.5	250	0.911	0.978	0.821	0.949	0.991	1.000	0.960	1.000
	500	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000
1	250	0.210	0.378	0.091	0.228	0.389	0.911	0.198	0.726
	500	0.658	0.794	0.444	0.658	0.882	0.999	0.695	0.989

m_3		$\text{bias}(\hat{\theta}_{max})$		$\text{sd}(\hat{\theta}_{max})$		$\text{bias}(\hat{\theta}_{min})$		$\text{sd}(\hat{\theta}_{min})$	
σ	n	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.2$
	100	0.0004	0.0003	0.009	0.008	−0.042	−0.049	0.025	0.020
0.1	250	0.0000	0.0001	0.004	0.003	−0.019	−0.043	0.022	0.020
	500	0.0000	0.0000	0.002	0.001	−0.006	−0.034	0.011	0.022
	100	0.0019	0.0060	0.025	0.026	−0.012	−0.022	0.028	0.029
0.5	250	0.0007	0.0027	0.011	0.012	−0.005	−0.015	0.013	0.020
	500	0.0002	0.0016	0.006	0.007	−0.003	−0.009	0.007	0.014
	100	−0.0043	0.0067	0.059	0.049	−0.012	−0.029	0.054	0.045
1	250	0.0019	0.0071	0.030	0.031	−0.012	−0.027	0.033	0.034
	500	0.0011	0.0047	0.017	0.019	−0.009	−0.024	0.021	0.029

Table 4.5: Simulation study as described in Section 4.3.3 for the regression function m_3.

4.4 Real data application: The Prague Temperature

This real data examples is based on a fixed, equidistant design, rather than a random regressor X. In this case we observe

$$Y_{n,i} = m\left(\tfrac{i}{n+1}\right) + \varepsilon_{n,i}, \quad \text{with independent and centered } \varepsilon_{n,i} \sim F_{n,i}\,, i = 1, \ldots, n.$$

If all the $F_{n,i}$, $i = 1, \ldots, n$, are equal and sufficiently smooth, then the limiting process is as in Remark 4.1 and (4.5), therefore Table 4.1 remains applicable, for n large enough. In the finite sample case, the only minor difference between fixed and random design is the domain of $V_{n,\alpha}$, which is I_0 in the above fixed design setting.

We consider the average annual temperatures in Prague from 1775 through 1989. In previous analyses it is found that the number of jump points is two or three, see Table 4.6. (Note that in Antonidiadis and Gijbels (2002) the data are studied only up to 1902.) From the HH-plot, the change-point recurring in the literature at around 1835 is easily detected. The plot, see Figure 4.6(b), is for the years in 1823–1835 below $-q_{0.2}(0.95)/\sqrt{n}$ and in 1911, 1914, 1928–1929

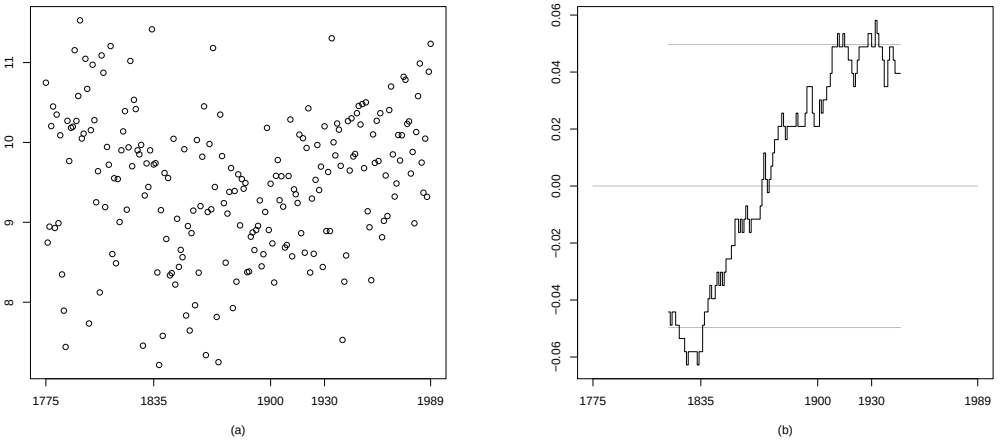

Figure 4.6: Average annual temperatures in Prague from 1775 to 1989; (a) scatter plot, (b) HH-plot for $\alpha = 0.2$.

methods	years of jump			
Horváth, Kokoszka and Steinebach (1999)		1835	1893	1927
Antonidiadis and Gijbels (2002)	1787	1837		–
Gijbels and Goderniaux (2004)	1786.5	1836.5		1942.5
HH-plot		1830		1932

Table 4.6: Jump points in the literature.

and 1932–1933 above $q_{0.2}(0.95)/\sqrt{n}$. Hence there is a drastic change between 1823 and 1835; according to the spikes we might locate it at 1830. After this jump down follows a strong increase of the regression curve with a possible jump up between 1911 and 1933. Using the ad hoc estimator, we can locate such a jump in the year 1932. This change is in accordance with changes found in the literature (1927 and 1942.5). The early jump found in the literature (at 1787) is difficult to assess since there are only 12 observations before 1787; this year is outside the domain of the HH-plot.

4.5 Proofs

Because α is fixed, it is henceforth dropped from the notation as a subscript. Set $\tilde{F}_n(u,y) := F_n(Q_1(u),y)$, $\tilde{F}_{n,1} := \tilde{F}_n(u,\infty)$, and denote its generalized inverse by $\tilde{F}_{n,1}^{-1}$. The corresponding standard Brownian bridge is $B_1(u) := B_F(Q_1(u),\infty)$. Because the marginals of $\tilde{F}(u,y)$, the identity and F_2, are uniformly continuous, $\tilde{F}$ itself is also uniformly continuous on $(0,1)\times\mathbb{R}$, and for all $(u,y)\in(0,1)\times\mathbb{R}$, since $\tilde{F}(u,\infty)=u$,

$$\tilde{F}'_x(u,y) \leq 1. \tag{4.7}$$

4.5.1 Lemmas

The proof of Theorem 4.1 is based on three lemmas.

Lemma 4.1. *Let* $S_n(u,y) := n^{\frac{1}{2}}\left(\tilde{F}\left(\tilde{F}_{n,1}^{-1}(u),y\right) - \tilde{F}\left(u,y\right)\right)$, *with* $0<u<1$, $y\in\mathbb{R}$, $n\in\mathbb{N}$. *With assumptions (A) and (B), on the probability space of*

(4.1),

$$\sup_{0<u<1}\sup_{y\in\mathbb{R}}\left|S_n(u,y)+\tilde{F}'_x(u,y)B_1(u)\right|\to 0 \quad a.s.\ as\ n\to\infty. \tag{4.8}$$

Proof With the mean-value theorem with u^* between u and $\tilde{F}^{-1}_{n,1}(u)$, we can write $S_n(u,y)=n^{\frac{1}{2}}\left(\tilde{F}^{-1}_{n,1}(u)-u\right)\tilde{F}'_x(u^*,y)$, and hence the left-hand side of (4.8) is bounded from above by

$$\begin{aligned}&\sup_{0<u<1}\left|n^{\frac{1}{2}}\left(\tilde{F}^{-1}_{n,1}(u)-u\right)+B_1(u)\right|\ \sup_{0<u<1}\sup_{y\in\mathbb{R}}\tilde{F}'_x(u^*,y)\\ &\quad+\sup_{0<u<1}|B_1(u)|\ \sup_{0<u<1}\sup_{y\in\mathbb{R}}\left|\tilde{F}'_x(u,y)-\tilde{F}'_x(u^*,y)\right|.\end{aligned} \tag{4.9}$$

From (4.1) we get that

$$\lim_{n\to\infty}\sup_{0<u<1}\left|n^{\frac{1}{2}}\left(\tilde{F}_{n,1}(u)-u\right)-B_1(u)\right|=0 \quad \text{a.s.},$$

from which it follows with the Vervaat (1972) lemma that

$$\lim_{n\to\infty}\sup_{0<u<1}\left|n^{\frac{1}{2}}\left(\tilde{F}^{-1}_{n,1}(u)-u\right)+B_1(u)\right|=0 \quad \text{a.s.} \tag{4.10}$$

Hence, with (4.10) and (4.7), the first term of (4.9) is equal to 0 as $n\to\infty$. With (4.10) and assumption (B), the second factor of the second term of (4.9) is going to 0 as $n\to\infty$, and because of the boundedness of B_1, we obtain (4.8). □

Let $\alpha\in(0,\frac{1}{2})$. For $x\in I_0$ and $y\in\mathbb{R}$, write

$$\begin{aligned}L_x(y) \ :=\ \tfrac{1}{2\alpha}\Big\{&B_F\big(Q_1(F_1(x)+\alpha),y\big)-B_F\big(Q_1(F_1(x)-\alpha),y\big)\\ &+\tilde{F}'_x\big(F_1(x)+\alpha,y\big)\ [B_F(x,\infty)\ -\ B_F\left(Q_1\left(F_1(x)+\alpha\right),\infty\right)]\\ &-\tilde{F}'_x\left(F_1(x)-\alpha,y\right)\ [B_F(x,\infty)\ -\ B_F\left(Q_1\left(F_1(x)-\alpha\right),\infty\right)]\Big\}.\end{aligned}$$

Note that

$$\sup_{(x,y)\in I_0\times\mathbb{R}} |L_x(y)| < \infty, \tag{4.11}$$

and that the functions $\{L_x : x \in I_0\}$ are uniformly equicontinuous.

Lemma 4.2. *Let $\alpha \in (0, \frac{1}{2})$. Under the assumptions (A) and (B), on the probability space of (4.1),*

$$\lim_{n\to\infty} \sup_{(x,y)\in I_n\times\mathbb{R}} \left| n^{\frac{1}{2}} \left[G_{n,x}(y) - G_x(y) \right] - L_x(y) \right| = 0 \quad a.s.$$

Proof First, rewrite $2\alpha n^{\frac{1}{2}} \left[G_{n,x}(y) - G_x(y) \right]$ in the following way:

$$\begin{aligned} & n^{\frac{1}{2}} \Big[\tilde{F}_n \left(\tilde{F}_{n,1}^{-1} \left(F_{n,1}(x) + \alpha \right), y \right) - \tilde{F} \left(F_1(x) + \alpha, y \right) \\ & \qquad - \tilde{F}_n \left(\tilde{F}_{n,1}^{-1} \left(F_{n,1}(x) - \tfrac{\lceil n\alpha \rceil}{n} \right), y \right) + \tilde{F} \left(F_1(x) - \alpha, y \right) \Big] \end{aligned} \tag{4.12}$$

$$= n^{\frac{1}{2}} \left[\tilde{F}_n \left(\tilde{F}_{n,1}^{-1} \left(F_{n,1}(x) + \alpha \right), y \right) - \tilde{F} \left(\tilde{F}_{n,1}^{-1} \left(F_{n,1}(x) + \alpha \right), y \right) \right] \tag{4.12a}$$

$$+ n^{\frac{1}{2}} \left[\tilde{F} \left(\tilde{F}_{n,1}^{-1} \left(F_{n,1}(x) + \alpha \right), y \right) - \tilde{F} \left(F_{n,1}(x) + \alpha, y \right) \right] \tag{4.12b}$$

$$+ n^{\frac{1}{2}} \left[\tilde{F} \left(F_{n,1}(x) + \alpha, y \right) - \tilde{F} \left(F_1(x) + \alpha, y \right) \right] \tag{4.12c}$$

$$\begin{aligned} - n^{\frac{1}{2}} \Big[& \tilde{F}_n \left(\tilde{F}_{n,1}^{-1} \left(F_{n,1}(x) - \tfrac{\lceil n\alpha \rceil}{n} \right), y \right) \\ & - \tilde{F} \left(\tilde{F}_{n,1}^{-1} \left(F_{n,1}(x) - \tfrac{\lceil n\alpha \rceil}{n} \right), y \right) \Big] \end{aligned} \tag{4.12d}$$

$$- n^{\frac{1}{2}} \left[\tilde{F} \left(\tilde{F}_{n,1}^{-1} \left(F_{n,1}(x) - \tfrac{\lceil n\alpha \rceil}{n} \right), y \right) - \tilde{F} \left(F_{n,1}(x) - \tfrac{\lceil n\alpha \rceil}{n}, y \right) \right] \tag{4.12e}$$

$$- n^{\frac{1}{2}} \left[\tilde{F} \left(F_{n,1}(x) - \tfrac{\lceil n\alpha \rceil}{n}, y \right) - \tilde{F} \left(F_1(x) - \alpha, y \right) \right]. \tag{4.12f}$$

From (4.1),

$$\lim_{n\to\infty} \sup_{(x,y)\in I_n\times\mathbb{R}} \Bigg| n^{\frac{1}{2}} \left[\tilde{F}_n \left(\tilde{F}_{n,1}^{-1} \left(F_{n,1}(x) + \alpha \right), y \right) - \tilde{F} \left(\tilde{F}_{n,1}^{-1} \left(F_{n,1}(x) + \alpha \right), y \right) \right] - B_F \left(Q_{n,1} \left(F_{n,1}(x) + \alpha \right), y \right) \Bigg| = 0 \quad \text{a.s.}$$

Because of the uniform continuity of $\tilde{F}$, with (4.10), (4.12a) converges almost surely for $n \to \infty$, uniformly in $x \in I_n$ and $y \in \mathbb{R}$, to $B_F \left(Q_1 \left(F_1(x) + \alpha \right), y \right)$. Similarly, the expression in (4.12d) converges almost surely for $n \to \infty$, uniformly in $x \in I_n$ and $y \in \mathbb{R}$, to $-B_F \left(Q_1 \left(F_1(x) - \alpha \right), y \right)$.

From Lemma 4.1, assumption (B), and the uniform continuity of B_1, we have that the expression in (4.12b) converges uniformly on $I_n \times \mathbb{R}$, as $n \to \infty$, to $-\tilde{F}'_x \left(F_1(x) + \alpha, y \right) \; B_1 \left(F_1(x) + \alpha \right)$ almost surely; similarly the expression in (4.12e) converges to $\tilde{F}'_x \left(F_1(x) - \alpha, y \right) \; B_1 \left(F_1(x) - \alpha \right)$ almost surely.

For the convergence of (4.12c) and (4.12f), a similar argument as for (4.12b) and (4.12e) holds. For the expression in (4.12f), with (4.1), (4.7) and the mean-value theorem,

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{(x,y)\in I_n\times\mathbb{R}} \Bigg| n^{\frac{1}{2}} \left[\tilde{F} \left(F_{n,1}(x) - \tfrac{\lceil n\alpha \rceil}{n}, y \right) - \tilde{F} \left(F_1(x) - \alpha, y \right) \right] \\
&\qquad - \tilde{F}'_x \left(F_1(x) - \alpha, y \right) \; B_F(x, \infty) \Bigg| \\
\leq \; &\lim_{n\to\infty} \sup_{(x,y)\in I_n\times\mathbb{R}} \Bigg| \frac{\tilde{F} \left(F_{n,1}(x) - \frac{\lceil n\alpha \rceil}{n}, y \right) - \tilde{F} \left(F_1(x) - \alpha, y \right)}{F_{n,1}(x) - F_1(x) + \alpha - \frac{\lceil n\alpha \rceil}{n}} \\
&\qquad - \tilde{F}'_x \left(F_1(x) - \alpha, y \right) \Bigg| \\
&\qquad \cdot \lim_{n\to\infty} \sup_{x\in\mathbb{R}} \left| n^{\frac{1}{2}} \left(F_{n,1}(x) - F_1(x) \right) \right| + \sup_{(x,y)\in I_0\times\mathbb{R}} \left| \tilde{F}'_x \left(F_1(x) - \alpha, y \right) \right| \\
&\qquad \cdot \lim_{n\to\infty} \sup_{x\in I_n} \left| n^{\frac{1}{2}} \left(F_{n,1}(x) - F_1(x) \right) - B_F(x, \infty) \right| \; = \; 0 \;\; \text{a.s.},
\end{aligned}$$

because of assumption (B). Similarly,

$$\lim_{n\to\infty} \sup_{(x,y)\in I_n\times\mathbb{R}} \Big| n^{\frac{1}{2}} \left[\tilde{F}\left(F_{n,1}(x)+\alpha, y\right) - \tilde{F}\left(F_1(x)+\alpha, y\right)\right] - \tilde{F}'_x(F_1(x)+\alpha, y)\, B_F(x,\infty) \Big| = 0 \quad \text{a.s. } \square$$

Lemma 4.3. *Under (A) and (B), we have for fixed $s \in (0,1)$, on the probability space of (4.1),*

$$\lim_{n\to\infty} \sup_{x\in I_n} |Q_{n,x}(s) - Q_x(s)| = 0 \quad a.s. \tag{4.13}$$

and

$$\lim_{n\to\infty} \sup_{x\in I_n} \left| n^{\frac{1}{2}} \left(Q_{n,x}(s) - Q_x(s)\right) + Q'_x(s) L_x\left(Q_x(s)\right)\right| = 0 \quad a.s. \tag{4.14}$$

Proof Since $G_x(\cdot)$ is increasing on $\mathbb{R}$,

$$\begin{aligned} |Q_{n,x}(s) - Q_x(s)| &= |Q_x\left(G_x\left(Q_{n,x}(s)\right)\right) - Q_x(s)| \\ &= \left| \frac{Q_x\left(G_x\left(Q_{n,x}(s)\right)\right) - Q_x(s)}{G_x\left(Q_{n,x}(s)\right) - s} \left(G_x\left(Q_{n,x}(s)\right) - s\right)\right|. \end{aligned} \tag{4.15}$$

Because Q_x is differentiable, the mean-value theorem yields that the right-hand side of (4.15) is equal to

$$|Q'_x(s_n^*)\left(G_x\left(Q_{n,x}(s)\right) - s\right)| = \left| \frac{1}{G'_x\left(Q_x(s_n^*)\right)} \left(G_x\left(Q_{n,x}(s)\right) - s\right)\right|, \tag{4.16}$$

for some s_n^* between $G_x\left(Q_{n,x}(s)\right)$ and s. From Lemma 4.2 it follows that, almost surely,

$$\begin{aligned} &\limsup_{n\to\infty} \sup_{x\in I_n} \left| n^{\frac{1}{2}} \left(G_x\left(Q_{n,x}(s)\right) - s\right) + L_x\left(Q_{n,x}(s)\right)\right| \\ &= \lim_{n\to\infty} \sup_{x\in I_n} \left| -n^{\frac{1}{2}} \left(G_{n,x}\left(Q_{n,x}(s)\right) - G_x\left(Q_{n,x}(s)\right)\right) + L_x\left(Q_{n,x}(s)\right)\right| = 0, \end{aligned} \tag{4.17}$$

and hence, since L_x bounded,

$$\lim_{n\to\infty} \sup_{x\in I_n} |G_x\left(Q_{n,x}(s)\right) - s| = 0 \quad \text{a.s.} \tag{4.18}$$

Note that $\inf_{x\in I_0} Q_1\left(F_1(x)+\alpha\right) - Q_1\left(F_1(x)-\alpha\right) > 0$ and, for $0 < s_1 < s < s_2 < 1$, uniformly in $x \in I_0$ almost surely for n large enough, $Q_x(s_n^*) \in [Q_2(2\alpha s_1), Q_2\left(1-2\alpha(1-s_2)\right)]$. Hence

$$\inf_{x\in I_n} G_x'\left(Q_x(s_n^*)\right) \;\geq\; \inf_{x\in I_0} \tfrac{1}{2\alpha} \int_{Q_1(F_1(x)-\alpha)}^{Q_1(F_1(x)+\alpha)} f\left(u, Q_x(s_n^*)\right) du \;>\; 0. \tag{4.19}$$

This proves (4.13).

From (4.15)–(4.17), with s_n^* between $G_x\left(Q_{n,x}(s)\right)$ and s, it follows immediately that the left-hand side of (4.14) is equal to

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{x\in I_n} |-Q_x'(s_n^*) L_x\left(Q_{n,x}(s)\right) + Q_x'(s) L_x\left(Q_x(s)\right)| \\
&= \limsup_{n\to\infty} \sup_{x\in I_n} |-Q_x'(s_n^*) \left[L_x\left(Q_{n,x}(s)\right)\right. \\
&\qquad\qquad \left. - L_x\left(Q_x(s)\right)\right] - L_x\left(Q_x(s)\right)\left[Q_x'(s_n^*) - Q_x'(s)\right]| \;.
\end{aligned} \tag{4.20}$$

Since the functions $\{L_x : x \in I_0\}$ are uniformly equicontinuous, with (4.13),

$$\lim_{n\to\infty} \sup_{x\in I_n} |L_x\left(Q_{n,x}(s)\right) - L_x\left(Q_x(s)\right)| = 0 \quad \text{a.s.} \tag{4.21}$$

As in (4.19), we have $\inf_{x\in I_n} G_x'\left(Q_x(s)\right) > 0$. With assumption (B), (4.13), and (4.19), it follows that, almost surely,

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{x\in I_n} |Q_x'(s_n^*) - Q_x'(s)| \\
&= \lim_{n\to\infty} \sup_{x\in I_n} \left| \frac{\tilde{F}_y'\left(F_1(x)+\alpha, Q_x(s)\right) - \tilde{F}_y'\left(F_1(x)+\alpha, Q_x(s_n^*)\right)}{2\alpha G_x'\left(Q_x(s_n^*)\right) G_x'\left(Q_x(s)\right)} \right. \\
&\qquad \left. + \frac{\tilde{F}_y'\left(F_1(x)-\alpha, Q_x(s_n^*)\right) - \tilde{F}_y'\left(F_1(x)-\alpha, Q_x(s)\right)}{2\alpha G_x'\left(Q_x(s_n^*)\right) G_x'\left(Q_x(s)\right)} \right| \;=\; 0.
\end{aligned} \tag{4.22}$$

Equations (4.11), (4.19), (4.21) and (4.22) give that (4.20) is equal to zero and hence equation (4.14) is proven. □

4.5.2 Proof of Theorem 4.1

For the proof of Theorem 4.1 it suffices to show that for $s \in \left\{\frac{1}{4}, \frac{3}{4}\right\}$, on the probability space of (4.1),

$$\sup_{x\in I_n}\Bigg| n^{\frac{1}{2}} \Bigg\{ \Big[F_n\left(x, Q_{n,x}\left(s\right)\right) - F_n\left(Q_{n,1}\left(F_{n,1}(x) - \tfrac{\lceil n\alpha\rceil}{n}\right), Q_{n,x}(s)\right)\Big] - \big[F\big(x, Q_x\left(s\right)\big) - F\big(Q_1\left(F_1(x)-\alpha\right), Q_x(s)\big)\big]\Bigg\} - B_s(x)\Bigg| \to 0 \quad \text{a.s. } as \;\; n\to\infty. \tag{4.23}$$

First, rewrite in the following way:

$$n^{\frac{1}{2}}\Bigg\{\Big[F_n\big(x, Q_{n,x}\left(s\right)\big) - F_n\left(Q_{n,1}\left(F_{n,1}(x) - \tfrac{\lceil n\alpha\rceil}{n}\right), Q_{n,x}(s)\right)\Big] - \Big[F\big(x, Q_x\left(s\right)\big) - F\big(Q_1\left(F_1(x)-\alpha\right), Q_x(s)\big)\Big]\Bigg\} \tag{4.24}$$

$$= n^{\frac{1}{2}}\big\{F_n\big(x, Q_{n,x}\left(s\right)\big) - F\big(x, Q_{n,x}\left(s\right)\big)\big\} \tag{4.24a}$$

$$+n^{\frac{1}{2}}\big\{F\big(x, Q_{n,x}\left(s\right)\big) - F\big(x, Q_x\left(s\right)\big)\big\} \tag{4.24b}$$

$$-n^{\frac{1}{2}}\Bigg\{\tilde{F}_n\left(\tilde{F}_{n,1}^{-1}\left(F_{n,1}(x) - \tfrac{\lceil n\alpha\rceil}{n}\right), Q_{n,x}(s)\right) - \tilde{F}\left(\tilde{F}_{n,1}^{-1}\left(F_{n,1}(x) - \tfrac{\lceil n\alpha\rceil}{n}\right), Q_{n,x}(s)\right)\Bigg\} \tag{4.24c}$$

$$-n^{\frac{1}{2}}\Bigg\{\tilde{F}\left(\tilde{F}_{n,1}^{-1}\left(F_{n,1}(x) - \tfrac{\lceil n\alpha\rceil}{n}\right), Q_{n,x}(s)\right) - \tilde{F}\left(F_{n,1}(x) - \tfrac{\lceil n\alpha\rceil}{n}, Q_{n,x}(s)\right)\Bigg\} \tag{4.24d}$$

$$-n^{\frac{1}{2}}\Bigg\{\tilde{F}\left(F_{n,1}(x) - \tfrac{\lceil n\alpha\rceil}{n}, Q_{n,x}(s)\right) - \tilde{F}\left(F_1(x)-\alpha, Q_{n,x}(s)\right)\Bigg\} \tag{4.24e}$$

$$-n^{\frac{1}{2}}\big\{\tilde{F}\left(F_1(x)-\alpha, Q_{n,x}(s)\right) - \tilde{F}\left(F_1(x)-\alpha, Q_x(s)\right)\big\}. \tag{4.24f}$$

The subtrahends (4.24c), (4.24d) and (4.24e) are equal to (4.12d), (4.12e) and (4.12f), respectively, evaluated at $y = Q_{n,x}(s)$. Therefore, with Lemma 4.3, uniformly in $x \in I_n$ for $n \to \infty$, with a similar reasoning as for (4.12d), (4.24c) converges to $-B_F\left(Q_1\left(F_1(x)-\alpha\right), Q_x(s)\right)$ almost surely. Analogous to (4.24c), (4.24a) converges uniformly in $x \in I_n$ for $n \to \infty$ almost surely to $B_F\left(x, Q_x(s)\right)$. Taking moreover (B) into account, it follows immediately from Lemma 4.1 and Lemma 4.3 that (4.24d) converges to $\tilde{F}'_x\left(F_1(x)-\alpha, Q_x(s)\right)$ $B_1\left(F_1(x)-\alpha\right)$ almost surely and (4.24e) converges to $-\tilde{F}'_x\left(F_1(x)-\alpha, Q_x(s)\right)$ $B_F\left(x,\infty\right)$ almost surely.

For (4.24b), we have

$$
\begin{aligned}
&\limsup_{n\to\infty}\sup_{x\in I_n}\Big|n^{\frac{1}{2}}\left[F\left(x, Q_{n,x}\left(s\right)\right)-F\left(x, Q_x\left(s\right)\right)\right] \\
&\quad + F'_y\left(x, Q_x(s)\right) Q'_x(s) L_x\left(Q_x(s)\right)\Big| \\
&= \limsup_{n\to\infty}\sup_{x\in I_n}\Bigg|\frac{F\left(x, Q_{n,x}(s)\right)-F\left(x, Q_x(s)\right)}{Q_{n,x}(s)-Q_x(s)}\, n^{\frac{1}{2}}\left(Q_{n,x}(s)-Q_x(s)\right) \\
&\quad + F'_y\left(x, Q_x(s)\right) Q'_x(s) L_x\left(Q_x(s)\right)\Bigg| \\
&\leq \limsup_{n\to\infty}\sup_{x\in I_n}\Bigg|\frac{F\left(x, Q_{n,x}(s)\right)-F\left(x, Q_x(s)\right)}{Q_{n,x}(s)-Q_x(s)}\ \Big[n^{\frac{1}{2}}\left(Q_{n,x}(s)-Q_x(s)\right) \\
&\quad + Q'_x(s) L_x\left(Q_x(s)\right)\Big]\Bigg| \ + \ \limsup_{n\to\infty}\sup_{x\in I_n}\Bigg|Q'_x(s) L_x\left(Q_x(s)\right) \\
&\quad \cdot \left[\frac{F\left(x, Q_{n,x}(s)\right)-F\left(x, Q_x(s)\right)}{Q_{n,x}(s)-Q_x(s)} - F'_y\left(x, Q_x(s)\right)\right]\Bigg| \,. \qquad (4.25)
\end{aligned}
$$

With assumption (B), Lemma 4.3, and the mean-value theorem, it follows that

$$
\lim_{n\to\infty}\sup_{x\in I_n}\left|\frac{F\left(x, Q_{n,x}(s)\right)-F\left(x, Q_x(s)\right)}{Q_{n,x}(s)-Q_x(s)} - F'_y\left(x, Q_x(s)\right)\right| = 0 \quad \text{a.s.}, \quad (4.26)
$$

and with assumption (A) we get

$$
\sup_{x\in I_n} F'_y\left(x, Q_x(s)\right) \leq \sup_{y\in\mathbb{R}} f_2(y) < \infty,
$$

which, with (4.26), bounds the first factor of the first term of the right-hand

side of (4.25). Lemma 4.3 yields that the second factor of this term is equal to zero almost surely. From (4.11), (4.19), and (4.26) it now follows directly that (4.25) is equal to zero almost surely.

Accordingly (4.24f) converges uniformly in $x \in I_n$ for $n \to \infty$, almost surely to

$$F'_y \left(Q_1 \left(F_1(x) - \alpha \right), Q_x(s) \right) \; Q'_x(s) \; L_x \left(Q_x(s) \right) \; . \qquad \square$$

Chapter 5

Testing for bivariate spherical symmetry

[Based on joint work with J.H.J. Einmahl, *working paper*.]

Abstract: An omnibus test for spherical symmetry in $\mathbb{R}^2$ is proposed, employing localized empirical likelihood. The thus obtained test statistic is distribution-free under the null hypothesis. The asymptotic null distribution is established and critical values for typical sample sizes, as well as the asymptotic ones, are presented. In a simulation study, the good performance of the test is demonstrated. Furthermore, a real data example is presented.

Key words: Asymptotic distribution, distribution-free, empirical likelihood, hypothesis test, spherical symmetry.

5.1 Introduction

Spherically symmetric distributions are an important class of distributions: They are a generalization of the multivariate standard normal distribution and include, amongst others, also multivariate Laplace and t distributions.

Furthermore, spherical symmetry is a distributional assumption which is associated with many statistical models, see Fang et al. (1990). For instance, only recently a relationship between L_1 spherical symmetry and Archimedean copulas was discovered in McNeil and Nešlehovà (2009). Another example is Laurent (1974), where univariate general linear models are considered with an error term that is spherically symmetric distributed. More applications of spherically symmetric distributions in statistics, such as in minimax estimation or stochastic processes, are discussed in Chmielewski (1981). For a general introduction to symmetry see Serfling (2006). Our focus is on spherical symmetry in $\mathbb{R}^2$.

There exist several approaches to test for spherical symmetry, cf. the survey paper Fang and Liang (1999) or Liang et al. (2008) for a good overview. An often used basis, that is also underlying this paper, is the stochastic representation: Let $X = (X_1, X_2)$ be a bivariate random vector. Define the radius $S := \sqrt{X_1^2 + X_2^2}$ and the direction $Z := X/S$. Then X is bivariate spherically symmetric (in the L_2-norm) if and only if S is independent of Z, and Z is uniformly distributed on the unit circle. Other nonparametric tests based on this stochastic representation include Smith (1977) and Baringhaus (1991), whereas the test proposed in Koltchinskii and Li (1998) uses multivariate distribution functions (df's) and a multivariate extension of quantile functions.

We will moreover use that uniform random variables on a circle which are projected on a tangent to that circle are Cauchy distributed on that tangent (see, amongst others, Szabłowski, 1998). In addition to the above definitions let $Y := X_2/X_1$. If Z is uniformly distributed on the unit circle it follows that $(1, Y)$, which is its projection on the tangent at $(1, 0)$, is standard Cauchy distributed. Since such a projection cannot distinguish between (X_1, X_2) and $(-X_1, -X_2)$, we project those (X_1, X_2) with $X_1 > 0$ on the line $x_1 = 1$, whereas the (X_1, X_2) with $X_1 < 0$ are projected on $x_1 = -1$. Denoting $\delta := \text{sign}(X_1)$, both $Y \,|\, \delta = -1$ and $Y \,|\, \delta = 1$ are then also standard Cauchy distributed.

We wish to test

$$H_0 : \ X \text{ is spherically symmetric around the origin}$$

on the basis of this stochastic representation, but, for the first time, the test is developed in an empirical likelihood framework.

The empirical likelihood method has the nice features which are known from parametric likelihood theory, but the data are used directly, i.e. in a nonparametric manner (see the monograph Owen, 2001). By localizing a functional equation, see Einmahl and McKeague (2003), we create an omnibus test for spherical symmetry. More precisely, a functional equation is 'split up' in infinitely many pointwise equations and then standard empirical likelihood theory is used to deal with these pointwise constraints. Finally the infinitely many likelihood ratios are considered simultaneously as a stochastic process and an integral of this stochastic process is taken.

For the test presented here, the center of the bivariate distribution has to be known. Nevertheless, the test could be extended to test for spherical symmetry around an unknown point by subtracting an estimated center from the data, for example Haldane's (1948) spatial median. In that case, the test statistic is not distribution-free under H_0 and hence the critical values are hard to find; therefore we restrict ourselves to a known center.

In Section 5.2, we derive the test statistic and present its limiting behavior under H_0. The test is consistent against all alternatives. In Section 5.3, critical values are computed, and in a simulation study we examine the performance of the test by power calculations for normal distributions and by a comparison to the test proposed in Koltchinskii and Li (1998). Furthermore, an application to a financial data set is presented. The proof of the main result is deferred to Section 5.4.

5.2 Main results

Let (S, Y, δ), as introduced in Section 5.1, have df F with marginals F_S, F_Y, and F_δ. Define the subdistribution functions by $F^-(s,y) := F(s,y,-1)$ and $F^+(s,y) := F(s,y,1) - F^-(s,y)$ and denote their marginals with $F_S^\pm$ and $F_Y^\pm$. Then the null hypothesis of spherical symmetry can be written as

$$H_0 : \; F^-(s,y) = F^+(s,y) = \tfrac{1}{2} F_S(s) G(y), \quad \text{for all } s \in \mathbb{R}^+,\, y \in \mathbb{R},$$

with G denoting the standard Cauchy df.

Remark 5.1. Within the localized empirical likelihood framework, a test based directly on (S, Z) can be constructed as well. Such a test has typically less power and is difficult to be generalized to higher dimensions because it is cumbersome to deal with the uniform distribution on the (hyper-)sphere. Therefore we use the transformation to (S, Y, δ), which, appropriately generalized to higher dimensions, leads to a Cauchy distributed Y on (hyper-)planes.

Consider n independent random variables $(X_{11}, X_{21}), \ldots, (X_{1n}, X_{2n})$ distributed as (X_1, X_2). Write (S_i, Y_i, δ_i), $i = 1, \ldots, n$, for the transformed random vectors and denote with F_n their empirical df. Define the nonparametric likelihood $L(\tilde{F}) = \prod_{i=1}^n \tilde{P}\left(\{(S_i, Y_i, \delta_i)\}\right)$, where $\tilde{P}$ is the probability measure corresponding to $\tilde{F}$. Furthermore, define for *fixed* $(s,y) \in \mathbb{R}^+ \times \mathbb{R}$, the *localized* empirical likelihood ratio

$$R(s,y) = \frac{\sup^\star\{L(\tilde{F})\}}{\sup\{L(\tilde{F})\}}, \tag{5.1}$$

where $\sup^\star$ is the supremum taken under the constraints given by H_0 and the corresponding marginal constraints:

$$\begin{aligned}
&\tilde{F}^-(s,y) = \tilde{F}_S^-(s) G(y), && \tilde{F}^+(s,y) = \tilde{F}_S^+(s) G(y),\\
&\tilde{F}_Y^-(y) = \tfrac{1}{2} G(y), && \tilde{F}_Y^+(y) = \tfrac{1}{2} G(y),\\
&\tilde{F}_S^-(s) = \tilde{F}_S^+(s) = \tfrac{1}{2} \tilde{F}_S(s), && \tilde{F}^-(\infty,\infty) = \tilde{F}^+(\infty,\infty) = \tfrac{1}{2},
\end{aligned}$$

and sup is the maximum over the unrestricted likelihood obtained at $\tilde{F} = F_n$, i.e., giving each observation mass $\frac{1}{n}$.

Define the bivariate empirical subdistribution functions

$$F_n^{\pm}(s,y) = \tfrac{1}{n}\sum_{i=1}^{n} 1_{[0,s]\times(-\infty,y]\times\{\pm 1\}}(S_i, Y_i, \delta_i),$$

and write $N := nF_n^{-}(\infty,\infty)$. Observe that N is the number of data points with $X_{1i} \le 0$.

Consider for (S_i, Y_i, δ_i), $i = 1, \ldots, n$, and either choice of sign, the regions

$$\begin{aligned} A_3^{\pm} &= [0,s]\times(y,\infty)\times\{\pm 1\}, & A_4^{\pm} &= (s,\infty)\times(y,\infty)\times\{\pm 1\},\\ A_1^{\pm} &= [0,s]\times(-\infty,y]\times\{\pm 1\}, & A_2^{\pm} &= (s,\infty)\times(-\infty,y]\times\{\pm 1\}. \end{aligned}$$

Denote with P_n the empirical measure corresponding to F_n. Let $F_{Sn}^{\pm}$ and $F_{Tn}^{\pm}$ denote the respective marginal df's of $F_n^{\pm}$. Observe that

$$\begin{aligned} P_n(A_3^{\pm}) &= F_{Sn}^{\pm}(s) - F_n^{\pm}(s,y), & P_n(A_4^{\pm}) &= F_n^{\pm}(\infty,\infty) - F_{Sn}^{\pm}(s) - F_{Sn}^{\pm}(y)\\ & & & \quad + F_n^{\pm}(s,y),\\ P_n(A_1^{\pm}) &= F_n^{\pm}(s,y), & P_n(A_2^{\pm}) &= F_{Sn}^{\pm}(y) - F_n^{\pm}(s,y). \end{aligned}$$

To maximize the numerator of (5.1), $\tilde{F}$ should put equal mass p_j^-, say, on each observation in A_j^- and mass p_j^+ on each observation in A_j^+, $j = 1, \ldots, 4$. Hence we need to maximize

$$\prod_{j=1}^{4} \left(p_j^-\right)^{nP_n(A_j^-)} \left(p_j^+\right)^{nP_n(A_j^+)}$$

under the constraints

$$\begin{aligned} nP_n(A_1^-)p_1^- &= \left(nP_n(A_1^-)p_1^- + nP_n(A_3^-)p_3^-\right) G(y),\\ nP_n(A_1^+)p_1^+ &= \left(nP_n(A_1^+)p_1^+ + nP_n(A_3^+)p_3^+\right) G(y),\\ nP_n(A_1^-)p_1^- + nP_n(A_2^-)p_2^- &= \tfrac{1}{2}G(y),\\ nP_n(A_1^+)p_1^+ + nP_n(A_2^+)p_2^+ &= \tfrac{1}{2}G(y), \end{aligned}$$

$$
\begin{aligned}
& nP_n(A_1^-)p_1^- + nP_n(A_3^-)p_3^- \;=\; nP_n(A_1^+)p_1^+ + nP_n(A_3^+)p_3^+, \\
& \sum_{j=1}^{4} p_j^- nP_n(A_j^-) = \tfrac{1}{2}, \\
& \sum_{j=1}^{4} p_j^+ nP_n(A_j^+) = \tfrac{1}{2}.
\end{aligned}
$$

This yields, for either choice of sign, the maximum empirical likelihood estimators

$$
\begin{aligned}
\hat{p}_3^{\pm} &= \tfrac{F_{Sn}(s)(1-G(y))}{2nP_n(A_3^{\pm})}, & \hat{p}_4^{\pm} &= \tfrac{(1-F_{Sn}(s))(1-G(y))}{2nP_n(A_4^{\pm})}, \\
\hat{p}_1^{\pm} &= \tfrac{F_{Sn}(s)G(y)}{2nP_n(A_1^{\pm})}, & \hat{p}_2^{\pm} &= \tfrac{(1-F_{Sn}(s))G(y)}{2nP_n(A_2^{\pm})}.
\end{aligned}
$$

Define, for either choice of sign,

$$
\begin{aligned}
\log R^{\pm}(s,y) = {} & nP_n(A_1^{\pm}) \log \frac{F_{Sn}(s)G(y)}{2P_n(A_1^{\pm})} \;+\; nP_n(A_2^{\pm}) \log \frac{(1-F_{Sn}(s))\,G(y)}{2P_n(A_2^{\pm})} \\
& + nP_n(A_3^{\pm}) \log \frac{F_{Sn}(s)(1-G(y))}{2P_n(A_3^{\pm})} \;+\; nP_n(A_4^{\pm}) \log \frac{(1-F_{Sn}(s))\,(1-G(y))}{2P_n(A_4^{\pm})},
\end{aligned} \tag{5.2}
$$

where $0 \log(a/0) = 0$, then we have

$$
\log R(s,y) = \log R^-(s,y) + \log R^+(s,y).
$$

Consider the test statistic

$$
T_n = -2 \int_{-\infty}^{\infty} \int_{0}^{\infty} \log R(s,y) dF_{Sn}(s) dG(y).
$$

Clearly, T_n is distribution-free; selected critical values are provided in Table 5.1.

We now consider the limiting distribution of T_n. In order to define the limiting random variable, we denote with W a standard Wiener process on $[0,1]^3$, i.e. a centered Gaussian process with $Cov\,(W(u,v,w), W(\tilde{u},\tilde{v},\tilde{w})) = (u \wedge \tilde{u})(v \wedge \tilde{v})(w \wedge \tilde{w})$, and with $B(u,v,w) = W(u,v,w) - uvwW(1,1,1)$ the

standard trivariate Brownian bridge. We also define $B^-(u,v) := B(u,v,\frac{1}{2})$ and $B^+(u,v) := B(u,v,1) - B^-(u,v)$. Observe $B^-(1,1) = -B^+(1,1)$. Furthermore, let, for either choice of sign, $W_0^\pm$ be a four-sided tied-down "half" Wiener process on $[0,1]^2$ defined by $W_0^\pm(u,v) := B^\pm(u,v) - vB^\pm(u,1) - uB^\pm(1,v) + uvB^\pm(1,1)$. Finally write

$$\begin{aligned} K(u,v) &= \frac{W_0^-(u,v)^2 + W_0^+(u,v)^2}{\frac{1}{2}u(1-u)v(1-v)} + 4B^-(1,1)^2 \\ &\quad + \frac{[B^-(u,1) - uB^-(1,1) - B^+(u,1) + uB^+(1,1)]^2}{u(1-u)} \\ &\quad + \frac{[B^-(1,v) - vB^-(1,1)]^2 + [B^+(1,v) - vB^+(1,1)]^2}{\frac{1}{2}v(1-v)}. \end{aligned}$$

Theorem 5.1. *Let F_S be continuous. Then, under H_0,*

$$T_n \xrightarrow{d} \int_0^1 \int_0^1 K(u,v)\, du\, dv.$$

The proof of the theorem is given in Section 5.4.

Note that for fixed s and y, under H_0,

$$-2\log R(s,y) \xrightarrow{d} K(F_S(s), G(y)) \stackrel{d}{=} \chi_6^2.$$

This is a special case of Owen's (2001) nonparametric version of the classical Wilks theorem.

5.3 Simulation results and real data example

Table 5.1 provides selected critical values for the proposed test statistic T_n. The values for $n = 50$, 100 and 200 are based on 100 000 samples in each case. For $n = \infty$, the quantiles of the limiting distribution are given, also based on 100 000 repetitions.

	Percentage points			
n	90%	95%	97.5%	99%
50	8.83	10.01	11.23	12.81
100	8.83	9.99	11.20	12.80
200	8.77	9.96	11.17	12.74
∞	8.61	9.83	11.02	12.66

Table 5.1: Critical values for the test for bivariate spherical symmetry.

To evaluate the power of the test (based on the critical values from Table 5.1), we regard data from a bivariate normal distribution with means 0, variances 1, and correlation ρ. The calculations, which are presented in Table 5.2, are based on 1000 replications. At the 5% significance level we see a high power for $\rho = 0.6$ ($n = 100$), and for $n = 200$, $\rho = 0.4$ is already well detected.

	$n = 100$				$n = 200$			
	Significance level							
ρ	10%	5%	2.5%	1%	10%	5%	2.5%	1%
0.1	0.10	0.05	0.03	0.01	0.17	0.08	0.04	0.02
0.2	0.20	0.11	0.07	0.03	0.37	0.22	0.12	0.06
0.3	0.38	0.24	0.14	0.06	0.68	0.54	0.41	0.24
0.4	0.61	0.46	0.31	0.17	0.94	0.86	0.78	0.60
0.5	0.86	0.75	0.62	0.44	1.00	0.99	0.98	0.93
0.6	0.98	0.95	0.90	0.75	1.00	1.00	1.00	1.00
0.7	1.00	1.00	0.99	0.97	1.00	1.00	1.00	1.00
0.8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 5.2: Power of the test for bivariate spherical symmetry for sample sizes $n = 100$ and $n = 200$.

Next, we compare the performance of our localized empirical likelihood (LEL-) test with the test proposed in Koltchinskii and Li (1998) (KL-test), see Table 5.3. It needs to be pointed out that the null hypothesis in Koltchinskii and Li (1998) is broader: There the center is unknown. Therefore the powers cannot be likened directly: A positive comparison for the LEL-test can be seen as an advise to use that test in case a center is given. We consider all the alternatives introduced in Koltchinskii and Li (1998):

$H_1^{(1)}$: $X_1 \sim Exp(1)$ and $X_2 \sim Exp(2)$, X_1 and X_2 independent, with $Exp(\lambda)$ the exponential distribution with mean $1/\lambda$;

$H_1^{(2)}$: $X_1 \sim N(0,1)$ and $X_2 \sim Exp(1)$, X_1 and X_2 independent;

$H_1^{(3)}$: Mixture (with parameter $1/2$) of two normal distributions with identity covariance matrices and with means $(-1.5, 0)$ and $(1.5, 0)$;

$H_1^{(4)}$: Uniform distribution on an equilateral triangle, centered at the origin.

Especially $H_1^{(1)}$ and $H_1^{(2)}$ are clearly visible as non-symmetric by the naked eye and should therefore lead to a high power. To center the data around the origin, we transform the data of $H_1^{(1)}$ and $H_1^{(2)}$ by subtracting the medians, hence we consider $(X_1 - \text{med}(X_1), X_2 - \text{med}(X_2))$. This is in line with Koltchinskii and Li (1998), where the empirical spatial median is chosen to estimate the center. The results are again based on 1000 repetitions of the LEL-test, whereas the results for the KL-test are taken from Koltchinskii and Li (1998, 100 repetitions). The LEL-test outperforms the KL-test in nearly every setting and typically performs even considerably better. For the alternative hypotheses $H_1^{(1)}$, $H_1^{(3)}$, and $H_1^{(4)}$, the LEL-test has for $n = 100$ already about the same power as the KL-test for $n = 200$. Only for $H_1^{(2)}$, $n = 100$, both tests have comparable power.

Finally we present a real data example. The bivariate data are the daily exchange rate log-returns of the Yen to the Dollar and the Pound to the Euro from January 2nd, 2009, to December 31st, 2009. The data set has size $n = 251$ and is available from `http://wrds-web.wharton.upenn.edu`, see Figure 5.1. The returns are known to be centered at the origin; this is affirmed by an estimated spatial median of $(-3.3 \cdot 10^{-5}, -3.0 \cdot 10^{-4})$. We want to test whether these data are spherically symmetric and find $T_n = 6.84$, which is clearly below the asymptotic critical value at the 10% significance level. Therefore we cannot reject the null hypothesis of spherical symmetry.

Distribution	n	Significance level					
		10%		5%		1%	
		LEL	KL	**LEL**	KL	**LEL**	KL
$H_1^{(1)}$	100	**1.00**	0.23	**1.00**	0.16	**1.00**	0.04
	200	**1.00**	0.94	**1.00**	0.86	**1.00**	0.55
$H_1^{(2)}$	100	**0.97**	0.92	**0.89**	0.90	**0.46**	0.52
	200	**1.00**	1.00	**1.00**	0.92	**1.00**	0.63
$H_1^{(3)}$	100	**0.93**	0.14	**0.83**	0.11	**0.39**	0.02
	200	**1.00**	0.92	**1.00**	0.83	**0.99**	0.44
$H_1^{(4)}$	100	**0.73**	0.47	**0.53**	0.21	**0.21**	0.07
	200	**0.99**	0.81	**0.97**	0.57	**0.78**	0.19

Table 5.3: Powers of the localized empirical likelihood (*LEL*) test and the test in Koltchinskii and Li (*KL*, 1998).

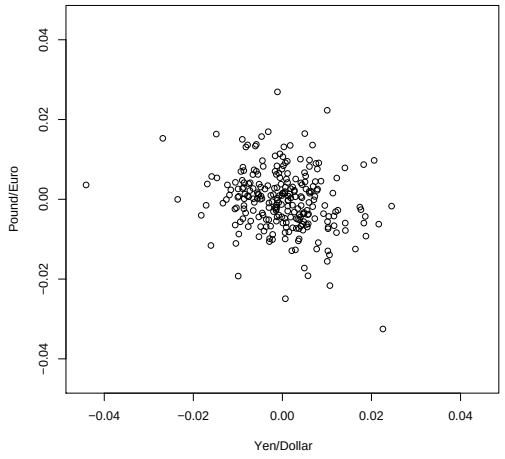

Figure 5.1: Daily exchange rate log-returns of Yen-Dollar and Pound-Euro, from January 2, 2009 to December 31, 2009.

5.4 Proof

Write Q_S, Q for the quantile functions corresponding to F_S, G, set $U_i = F_S(S_i)$ and $V_i = G(Y_i)$, and let Γ_n be the empirical df of the $(U_i, V_i, F_\delta(\delta_i))$ and Γ_{Sn}, Γ_{Yn} and $\Gamma_{\delta n}$ the corresponding marginals. Furthermore, write $\Gamma_n^-(u,v) := \Gamma_n(u,v,\frac{1}{2})$, hence Γ_n^- is the empirical subdistribution function of the (U_i, V_i),

for which $\delta_i = -1$, with marginals Γ^-_{Sn} and Γ^-_{Yn}, and note that $\Gamma^-_n(1,1) = \frac{N}{n}$. Define Γ^+_n similarly.

Let $0 < \varepsilon \leq \frac{1}{2}$. It suffices to show that, as $n \to \infty$,

$$\begin{aligned} T_{1n} &= -2\int_{Q(\varepsilon)}^{Q(1-\varepsilon)} \int_{Q_S(\varepsilon)}^{Q_S(1-\varepsilon)} \log R(s,y)\, dF_{Sn}(u)\, dG(y) \\ &\xrightarrow{d} \int_{\varepsilon}^{1-\varepsilon} \int_{\varepsilon}^{1-\varepsilon} K(u,v)\, du\, dv, \end{aligned} \tag{5.3}$$

and

$$T_{2n} = T_n - T_{1n} = O_P(\sqrt{\varepsilon}) \tag{5.4}$$

uniformly in ε; see Billingsley (1968, Theorem 4.2).

First, consider T_{1n} and decompose it further to

$$\begin{aligned} T_{1n} &= -2\int_{\varepsilon}^{1-\varepsilon} \int_{\varepsilon}^{1-\varepsilon} \log R^-(Q_S(u), Q(v))\, d\Gamma_{Sn}(u) dv \\ &\quad -2\int_{\varepsilon}^{1-\varepsilon} \int_{\varepsilon}^{1-\varepsilon} \log R^+(Q_S(u), Q(v))\, d\Gamma_{Sn}(u)\, dv \;=:\; T^-_{1n} + T^+_{1n}. \end{aligned}$$

Because of symmetry, we will first only consider T^-_{1n}. From (5.2), applying a Taylor expansion of $\log(1+x)$, it follows that, uniformly in $s \in [Q_S(\varepsilon), Q_S(1-\varepsilon)]$ and $y \in [Q(\varepsilon), Q(1-\varepsilon)]$,

$$\begin{aligned} \log R^-(s,y) &= \frac{n}{2} - N - \frac{n}{8}\left[\frac{\left(F_{Sn}(s)G(y) - 2P_n(A_1^-)\right)^2}{P_n(A_1^-)}\right. \\ &\quad + \frac{\left(\left(1 - F_{Sn}(s)\right)G(y) - 2P_n(A_2^-)\right)^2}{P_n(A_2^-)} + \frac{\left(F_{Sn}(s)\left(1 - G(y)\right) - 2P_n(A_3^-)\right)^2}{P_n(A_3^-)} \\ &\quad \left. + \frac{\left(\left(1 - F_{Sn}(s)\right)\left(1 - G(y)\right) - 2P_n(A_4^-)\right)^2}{P_n(A_4^-)}\right] + o_P(1) \\ &= \frac{n}{2} - N - \frac{F^-_{Sn}(s)}{8P_n(A_1^-)P_n(A_3^-)}\left[\sqrt{n}\left(F_{Sn}(s)G(y) - 2P_n(A_1^-)\right)\right]^2 \end{aligned}$$

$$
\begin{aligned}
&- \frac{\frac{N}{n} - F_{Sn}^{-}(s)}{8P_n(A_2^-)P_n(A_4^-)} \left[\sqrt{n}\left((1 - F_{Sn}(s))\, G(y) - 2P_n(A_2^-)\right)\right]^2 \\
&- \frac{\frac{N}{n} - F_{Tn}^{-}(y)}{8P_n(A_3^-)P_n(A_4^-)} \left[\sqrt{n}\left(F_{Sn}^{+}(s) - F_{Sn}^{-}(s)\right)\right]^2 \; - \; \frac{\left[\sqrt{n}\left(1 - 2\frac{N}{n}\right)\right]^2}{8P_n(A_4^-)} \\
&+ 2\frac{\sqrt{n}\left(1 - 2\frac{N}{n}\right)}{8P_n(A_4^-)} \left[\sqrt{n}\left(F_{Sn}^{+}(s) - F_{Sn}^{-}(s)\right) + \sqrt{n}\left((1 - F_{Sn}(s))\, G(y)\right.\right. \\
&\left.\left. - 2P_n(A_2^-)\right)\right] \; + \; 2\frac{\sqrt{n}\left(F_{Sn}^{+}(s) - F_{Sn}^{-}(s)\right)}{8P_n(A_3^-)} \sqrt{n}\left(F_{Sn}(s)G(y) - 2P_n(A_1^-)\right) \\
&- 2\frac{\sqrt{n}\left(F_{Sn}^{+}(s) - F_{Sn}^{-}(s)\right)}{8P_n(A_4^-)} \sqrt{n}\left((1 - F_{Sn}(s))\, G(y) - 2P_n(A_2^-)\right) \; + \; o_P(1).
\end{aligned}
$$

Observe that

$$
\begin{aligned}
&\sqrt{n}\left(F_{Sn}(s)G(y) - 2P_n(A_1^-)\right) \\
&\qquad = \; \sqrt{n}\left(F_{Sn}(s) - F_S(s)\right) G(y) \; - \; 2\sqrt{n}\left(P_n(A_1^-) - \tfrac{1}{2}F_S(s)G(y)\right),
\end{aligned}
$$

$$
\begin{aligned}
&\sqrt{n}\left((1 - F_{Sn}(s))\, G(y) - 2P_n(A_2^-)\right) \\
&\qquad = -2\sqrt{n}\left(F_{Tn}^{-}(y) - \tfrac{1}{2}G(y)\right) - \sqrt{n}\left(F_{Sn}(s)G(y) - 2P_n(A_1^-)\right),
\end{aligned}
$$

and

$$
\sqrt{n}\left(F_{Sn}^{+}(s) - F_{Sn}^{-}(s)\right) = \sqrt{n}\left(F_{Sn}^{+}(s) - F_S^{+}(s)\right) - \sqrt{n}\left(F_{Sn}^{-}(s) - F_S^{-}(s)\right).
$$

It follows from the Glivenko-Cantelli theorem that

$$
\frac{N}{n} \xrightarrow{P} \frac{1}{2},
$$

$$
\sup_{\substack{Q_S(\varepsilon) \le s \le Q_S(1-\varepsilon) \\ Q(\varepsilon) \le y \le Q(1-\varepsilon)}} \left| \frac{F_{Sn}^{-}(s)}{8P_n(A_1^-)P_n(A_3^-)} - \frac{1}{4F_S(s)G(y)\,(1 - G(y))} \right| \; = \; o_P(1),
$$

$$
\sup_{\substack{Q_S(\varepsilon) \le s \le Q_S(1-\varepsilon) \\ Q(\varepsilon) \le y \le Q(1-\varepsilon)}} \left| \frac{\frac{N}{n} - F_{Sn}^{-}(s)}{8P_n(A_2^-)P_n(A_4^-)} - \frac{1}{4\,(1 - F_S(s))\,G(y)\,(1 - G(y))} \right| \; = \; o_P(1),
$$

$$
\sup_{\substack{Q_S(\varepsilon) \le s \le Q_S(1-\varepsilon) \\ Q(\varepsilon) \le y \le Q(1-\varepsilon)}} \left| \frac{\frac{N}{n} - F_{Tn}^{-}(y)}{8P_n(A_3^-)P_n(A_4^-)} - \frac{1}{4F_S(s)\,(1 - F_S(s))\,(1 - G(y))} \right| \; = \; o_P(1),
$$

$$\sup_{\substack{Q_S(\varepsilon)\le s\le Q_S(1-\varepsilon)\\ Q(\varepsilon)\le y\le Q(1-\varepsilon)}}\left|\frac{1}{8P_n(A_3^-)}-\frac{1}{4F_S(s)\,(1-G(y))}\right| \;=\; o_P(1),$$

$$\sup_{\substack{Q_S(\varepsilon)\le s\le Q_S(1-\varepsilon)\\ Q(\varepsilon)\le y\le Q(1-\varepsilon)}}\left|\frac{1}{8P_n(A_4^-)}-\frac{1}{4\,(1-F_S(s))\,(1-G(y))}\right| \;=\; o_P(1). \tag{5.5}$$

Writing $\alpha_n^-(u,v) := \sqrt{n}\left(\Gamma_n^-(u,v)-\frac{1}{2}uv\right)$, $\alpha_n^+(u,v) := \sqrt{n}\left(\Gamma_n^+(u,v)-\frac{1}{2}uv\right)$, and $\alpha_n(u,v) := \alpha_n^-(u,v)+\alpha_n^+(u,v)$, we have, using (5.5) uniformly for $\varepsilon \le u,v \le 1-\varepsilon$,

$$\begin{aligned}
&\log R^-(Q_S(u),Q(v))\\
&= \frac{n}{2}-N \;-\; \frac{\left[\sqrt{n}\,(\Gamma_{Sn}(u)-u)\,v-2\sqrt{n}\left(\Gamma_n^-(u,v)-\frac{1}{2}uv\right)\right]^2}{4uv(1-v)}\\
&\quad-\frac{\left[2\sqrt{n}\left(\Gamma_{Yn}^-(v)-\frac{1}{2}v\right)+\sqrt{n}\,(\Gamma_{Sn}(u)-u)\,v-2\sqrt{n}\left(\Gamma_n^-(u,v)-\frac{1}{2}uv\right)\right]^2}{4(1-u)v(1-v)}\\
&\quad-\frac{\left[\sqrt{n}\left(\Gamma_{Sn}^+(u)-\frac{1}{2}u\right)-\sqrt{n}\left(\Gamma_{Sn}^-(u)-\frac{1}{2}u\right)\right]^2}{4u(1-u)(1-v)} \;-\; \frac{\left[\sqrt{n}\left(\Gamma_n^-(1,1)-\frac{1}{2}\right)\right]^2}{(1-u)(1-v)}\\
&\quad-\frac{\sqrt{n}\left(\Gamma_n^-(1,1)-\frac{1}{2}\right)}{(1-u)(1-v)}\left[\sqrt{n}\left(\Gamma_{Sn}^+(u)-\frac{1}{2}u\right)-\sqrt{n}\left(\Gamma_{Sn}^-(u)-\frac{1}{2}u\right)\right.\\
&\qquad\left.-2\sqrt{n}\left(\Gamma_{Yn}^-(v)-\frac{1}{2}v\right)-\sqrt{n}\,(\Gamma_{Sn}(u)-u)\,v+2\sqrt{n}\left(\Gamma_n^-(u,v)-\frac{1}{2}uv\right)\right]\\
&\quad+\frac{\sqrt{n}\left(\Gamma_{Sn}^+(u)-\frac{1}{2}u\right)-\sqrt{n}\left(\Gamma_{Sn}^-(u)-\frac{1}{2}u\right)}{2u(1-u)(1-v)}\left[\sqrt{n}\,(\Gamma_{Sn}(u)-u)\,v\right.\\
&\qquad\left.-2\sqrt{n}\left(\Gamma_n^-(u,v)-\frac{1}{2}uv\right)+2u\sqrt{n}\left(\Gamma_{Yn}^-(v)-\frac{1}{2}v\right)\right]+\;o_P(1)\\
&= \frac{n}{2}-N-\frac{\left[v\alpha_n(u,1)-2\alpha_n^-(u,v)\right]^2}{4uv(1-v)}-\frac{\left[2\alpha_n^-(1,v)+v\alpha_n(u,1)-2\alpha_n^-(u,v)\right]^2}{4(1-u)v(1-v)}\\
&\quad-\frac{\left[\alpha_n^+(u,1)-\alpha_n^-(u,1)\right]^2}{4u(1-u)(1-v)}-\frac{\alpha_n^-(1,1)^2}{(1-u)(1-v)}\\
&\quad-\frac{\alpha_n^-(1,1)}{(1-u)(1-v)}\left[\alpha_n^+(u,1)-\alpha_n^-(u,1)-2\alpha_n^-(1,v)-v\alpha_n(u,1)+2\alpha_n^-(u,v)\right]\\
&\quad+\frac{\alpha_n^+(u,1)-\alpha_n^-(u,1)}{2u(1-u)(1-v)}\left[v\alpha_n(u,1)+2u\alpha_n^-(1,v)-2\alpha_n^-(u,v)\right]\;+\;o_P(1).
\end{aligned}\tag{5.6}$$

Applying

$$-\frac{\left[v\alpha_n(u,1)-2\alpha_n^-(u,v)\right]^2}{4uv(1-v)}-\frac{\left[2\alpha_n^-(1,v)+v\alpha_n(u,1)-2\alpha_n^-(u,v)\right]^2}{4(1-u)v(1-v)}$$
$$=-\frac{\left[v\alpha_n(u,1)-2\alpha_n^-(u,v)\right]^2}{4u(1-u)v(1-v)}-\frac{\alpha_n^-(1,v)}{(1-u)v(1-v)}\left[v\alpha_n(u,1)-2\alpha_n^-(u,v)\right]$$
$$-\frac{\alpha_n^-(1,v)^2}{(1-u)v(1-v)},$$

$$-\frac{\left[v\alpha_n(u,1)-2\alpha_n^-(u,v)\right]^2}{4uv(1-v)(1-u)}-\frac{\left[\alpha_n^+(u,1)-\alpha_n^-(u,1)\right]^2}{4u(1-u)(1-v)}=-\frac{\left[\alpha_n^+(u,1)-\alpha_n^-(u,1)\right]^2}{4u(1-u)}$$
$$-\frac{\left[v\alpha_n^-(u,1)-\alpha_n^-(u,v)\right]^2}{u(1-u)v(1-v)}-\frac{\alpha_n^+(u,1)-\alpha_n^-(u,1)}{2u(1-u)(1-v)}\left[v\alpha_n(u,1)-2\alpha_n^-(u,v)\right],$$

$$-\frac{\left[v\alpha_n^-(u,1)-\alpha_n^-(u,v)\right]^2}{uv(1-v)(1-u)}-\frac{\alpha_n^-(1,v)^2}{(1-u)v(1-v)}=2\frac{\alpha_n^-(1,v)}{(1-u)v(1-v)}\left[v\alpha_n^-(u,1)\right.$$
$$\left.-\alpha_n^-(u,v)\right]-\frac{\left[v\alpha_n^-(u,1)-\alpha_n^-(u,v)+u\alpha_n^-(1,v)\right]^2}{u(1-u)v(1-v)}-\frac{\alpha_n^-(1,v)^2}{v(1-v)},$$

$$-\frac{\left[v\alpha_n^-(u,1)-\alpha_n^-(u,v)+u\alpha_n^-(1,v)\right]^2}{uv(1-v)(1-u)}-\frac{\alpha_n^-(1,1)^2}{(1-u)(1-v)}$$
$$=-\frac{\left[v\alpha_n^-(u,1)+u\alpha_n^-(1,v)-\alpha_n^-(u,v)-uv\alpha_n^-(1,1)\right]^2}{uv(1-v)(1-u)}-(1-uv)$$
$$\cdot\frac{\alpha_n^-(1,1)^2}{(1-u)(1-v)}-2\frac{\alpha_n^-(1,1)}{(1-u)(1-v)}\left[v\alpha_n^-(u,1)+u\alpha_n^-(1,v)-\alpha_n^-(u,v)\right],$$

$$-\frac{\alpha_n^-(1,v)^2}{v(1-v)}+\frac{uv\alpha_n^-(1,1)^2}{(1-u)(1-v)}=-\frac{\left[\alpha_n^-(1,v)-v\alpha_n^-(1,1)\right]^2}{v(1-v)}$$
$$+2\frac{\alpha_n^-(1,v)}{1-v}\alpha_n^-(1,1)+\frac{v\alpha_n^-(1,1)^2}{(1-u)(1-v)},$$

$$-\frac{\left[\alpha_n^+(u,1)-\alpha_n^-(u,1)\right]^2}{4u(1-u)}-\frac{\alpha_n^-(1,1)}{1-u}\left[\alpha_n^+(u,1)-\alpha_n^-(u,1)\right]$$
$$=-\frac{\left[\alpha_n^+(u,1)-\alpha_n^-(u,1)+2u\alpha_n^-(1,1)\right]^2}{4u(1-u)}+\frac{u\alpha_n^-(1,1)^2}{1-u},$$

to the right-hand side of (5.6) yields

$$\begin{aligned}
&\log R^-(Q_S(u), Q(v)) \\
&= \frac{n}{2} - N - \frac{\left[v\alpha_n^-(u,1) + u\alpha_n^-(1,v) - \alpha_n^-(u,v) - uv\alpha_n^-(1,1)\right]^2}{uv(1-v)(1-u)} - \alpha_n^-(1,1)^2 \\
&\quad - \frac{\left[\alpha_n^-(1,v) - v\alpha_n^-(1,1)\right]^2}{v(1-v)} - \frac{\left[\alpha_n^+(u,1) - \alpha_n^-(u,1) + 2u\alpha^-(1,1)\right]^2}{4u(1-u)} + o_P(1).
\end{aligned}$$

Because of symmetry we obtain a similar expression for $\log R^+(Q_S(u), Q(v))$. Hence we find

$$\begin{aligned}
&-2\log R(Q_S(u), Q(v)) \\
&= \frac{\left[\alpha_n^-(u,v) - v\alpha_n^-(u,1) - u\alpha_n^-(1,v) + uv\alpha_n^-(1,1)\right]^2}{\frac{1}{2}u(1-u)v(1-v)} \\
&\quad + \frac{\left[\alpha_n^+(u,v) - v\alpha_n^+(u,1) - u\alpha_n^+(1,v) + uv\alpha_n^+(1,1)\right]^2}{\frac{1}{2}u(1-u)v(1-v)} \\
&\quad + \frac{\left[\alpha_n^-(1,v) - v\alpha_n^-(1,1)\right]^2}{\frac{1}{2}v(1-v)} + \frac{\left[\alpha_n^+(1,v) - v\alpha_n^+(1,1)\right]^2}{\frac{1}{2}v(1-v)} + \frac{\alpha_n^-(1,1)^2}{\frac{1}{4}} \\
&\quad + \frac{\left[\alpha_n^+(u,1) - u\alpha_n^+(1,1) - \alpha_n^-(u,1) + u\alpha_n^-(1,1)\right]^2}{u(1-u)} + o_P(1).
\end{aligned}$$

Standard empirical process theory and the Skorohod construction (but keeping the same notation), yield, for either choice of sign,

$$\sup_{0\le u,v\le 1} \left|\alpha_n^{\pm}(u,v) - B^{\pm}(u,v)\right| \to 0 \quad \text{a.s.}$$

Hence T_{1n} can be replaced by

$$\int_{\varepsilon}^{1-\varepsilon}\int_{\varepsilon}^{1-\varepsilon} K(u,v)\,dv\,d\Gamma_{Sn}(u).$$

Because the integrand is uniformly continuous, this implies (5.3) by the Helly-Bray theorem.

To show (5.4), we only consider integration over the L-shaped region

$$C_\varepsilon = \left\{(u,v) \in (0,1)^2 :\ 0 < u \le \varepsilon,\ 0 < v \le \tfrac{1}{2} \ \text{ or } \ 0 < u \le \tfrac{1}{2},\ 0 < v \le \varepsilon\right\},$$

because of symmetry arguments. Consider the following five regions

$$\begin{aligned}
C_{\varepsilon,1,1} &= \left\{(u,v) \in (0,1)^2 :\ 0 < u \le n^{-3/5},\ n^{-3/8} \le v \le \tfrac{1}{2}\right\},\\
C_{\varepsilon,1,2} &= \left\{(u,v) \in (0,1)^2 : n^{-3/8} \le u \le \tfrac{1}{2},\ 0 < v \le n^{-3/5}\right\},\\
C_{\varepsilon,2} &= \left\{(u,v) \in (0, n^{-3/8}]^2\right\},\\
C_{\varepsilon,3,1} &= \left\{(u,v) \in (0,1)^2 :\ n^{-3/5} < u \le \varepsilon,\ n^{-3/8} \le v \le \tfrac{1}{2}\right\}\\
C_{\varepsilon,3,2} &= \left\{(u,v) \in (0,1)^2 :\ n^{-3/8} \le u \le \tfrac{1}{2},\ n^{-3/5} < v \le \varepsilon\right\},
\end{aligned}$$

which cover C_ε. We will use the following bound: For any $\eta > 0$ there exists a positive constant M_η, such that

$$\mathbb{P}\left(\Gamma_{Sn}(u) \ \le \ uM_\eta,\, \Gamma_{Yn}(u) \ \le \ uM_\eta, \ \text{ for all } 0 \le u \le 1\right) \ > \ 1 - \eta\,, \qquad (5.7)$$

see Shorack and Wellner (1986, p. 419).

For $C_{\varepsilon,1,1}$, $C_{\varepsilon,1,2}$, and $C_{\varepsilon,2}$ we only consider $\log R^-$, $\log R^+$ is treated similarly. We regard the four terms of (5.2) separately. For $C_{\varepsilon,1,1}$ and $C_{\varepsilon,2}$ we get, with (5.7) and if $P_n(A_j^\pm) \ge \frac{1}{n}$, $j = 1, \ldots, 4$, with probability $1 - \eta$,

$$\begin{aligned}
&\left| n\Gamma_n^-(u,v) \log \frac{\Gamma_{Sn}(u)v}{2\Gamma_n^-(u,v)} \right| \ \le \ n\Gamma_{Sn}(u) \log\left(\frac{v}{\frac{2}{n}} \ \vee \ \frac{2\Gamma_{Yn}(v)}{\Gamma_{Sn}(u)v} \right)\\
&\le \ M_\eta u n \log\left(n \ \vee \ \frac{2M_\eta}{\frac{1}{n}} \right) \ \le \ M_\eta u n \log(2M_\eta n),
\end{aligned}$$

and

$$\begin{aligned}
&\left| n\left(\Gamma_{Sn}^-(u) - \Gamma_n^-(u,v)\right) \log \frac{\Gamma_{Sn}(u)(1-v)}{2\left(\Gamma_{Sn}^-(u) - \Gamma_n^-(u,v)\right)} \right|\\
&\le \ n\Gamma_{Sn}(u) \log\left(n \ \vee \ \frac{2(1-\Gamma_{Yn}(v))}{\Gamma_{Sn}(u)(1-v)} \right) \ \le \ M_\eta u\, n \log(4n),
\end{aligned}$$

and, with $|\log(1+x)| \le 2|x|$ for $x \ge -0.5$, with probability $1-\eta$,

$$\begin{aligned}
&\left| n\left(\tfrac{N}{n} - \Gamma^-_{Sn}(u) - \Gamma^-_{Yn}(v) + \Gamma^-_n(u,v)\right) \log \frac{(1-\Gamma_{Sn}(u))\,(1-v)}{2\left(\tfrac{N}{n} - \Gamma^-_{Sn}(u) - \Gamma^-_{Yn}(v) + \Gamma^-_n(u,v)\right)} \right| \\
&\le n\left|(1-\Gamma_{Sn}(u))\,(1-v) - 2\tfrac{N}{n} + 2\Gamma^-_{Sn}(u) + 2\Gamma^-_{Yn}(v) - 2\Gamma^-_n(u,v)\right| \\
&\le n\left|\Gamma_{Sn}(u)v - 2\Gamma^-_n(u,v)\right| \;+\; n\left|2\Gamma^-_{Yn}(v) - v\right| + n\left|\Gamma_{Sn}(u) - 2\Gamma^-_{Sn}(u)\right| \\
&\quad + n\left|1 - 2\tfrac{N}{n}\right| \\
&\le n\left(\Gamma_{Sn}(u) + 2\Gamma^-_n(u,v)\right) \;+\; n\left|2\Gamma^-_{Yn}(v) - v\right| + n\left(\Gamma_{Sn}(u) + 2\Gamma^-_{Sn}(u)\right) \\
&\quad + n\left|1 - 2\tfrac{N}{n}\right| \\
&\le 6nM_\eta u \;+\; 2n\left|\Gamma^-_{Yn}(v) - \tfrac{1}{2}v\right| \;+\; 2n\left|\Gamma^-_n(1,1) - \tfrac{1}{2}\right| \\
&= \; 6nM_\eta u \;+\; 2n^{1/2}\left|\alpha^-_n(1,v)\right| + 2n^{1/2}\left|\alpha^-_n(1,1)\right|. \qquad (5.8)
\end{aligned}$$

Furthermore, for $C_{\varepsilon,2}$, we have, with probability $1-\eta$,

$$\left| n\left(\Gamma^-_{Yn}(v) - \Gamma^-_n(u,v)\right) \log \frac{(1-\Gamma_{Sn}(u))\,v}{2\left(\Gamma^-_{Yn}(v) - \Gamma^-_n(u,v)\right)} \right| \;\le\; M_\eta v\, n \log(2M_\eta n),$$

and for $C_{\varepsilon,1,1}$, employing the Taylor expansion as in (5.8), with probability $1-\eta$,

$$\begin{aligned}
&\left| n\left(\Gamma^-_{Yn}(v) - \Gamma^-_n(u,v)\right) \log \frac{(1-\Gamma_{Sn}(u))\,v}{2\left(\Gamma^-_{Yn}(v) - \Gamma^-_n(u,v)\right)} \right| \\
&\le n\left|(1-\Gamma_{Sn}(u))\,v - 2\Gamma^-_{Yn}(v) + 2\Gamma^-_n(u,v)\right| \;\le 3nM_\eta u \;+\; 2n\left|\Gamma^-_{Yn}(v) - \tfrac{1}{2}v\right| \\
&= \; 3nM_\eta u \;+\; 2n^{1/2}\left|\alpha^-_n(1,v)\right|.
\end{aligned}$$

Combining the above, we have that with probability $1-2\eta$

$$\begin{aligned}
&\iint_{C_{\varepsilon,1,1}} \left|\log R^-(Q_S(u), Q(v))\right| \; d\Gamma_{Sn}(u)\, dv \;\le\; \iint_{C_{\varepsilon,1,1}} M_\eta u n \left(\log(2M_\eta n)\right. \\
&\quad \left. +\; \log(4n)\right) + 9nM_\eta u + 4n^{1/2}\left|\alpha^-_n(1,v)\right| + 2n^{1/2}\left|\alpha^-_n(1,1)\right| \; d\Gamma_{Sn}(u)\, dv \\
&\le \; \left(2M_\eta n^{2/5}\log(4M_\eta n) + 9M_\eta n^{2/5} + 4n^{1/2}\sup_{0\le v\le 1}\left|\alpha^-_n(1,v)\right| + 2n^{1/2}\left|\alpha^-_n(1,1)\right|\right) \\
&\quad \cdot \int_{n^{-3/8}}^{1/2}\int_0^{n^{-3/5}} d\Gamma_{Sn}(u)\, dv \;\to\; 0,
\end{aligned}$$

and

$$\begin{aligned}
&\iint_{C_{\varepsilon,2}} \left|\log R^-(Q_S(u),Q(v))\right| \, d\Gamma_{Sn}(u)\,dv \;\le\; \iint_{C_{\varepsilon,2}} (M_\eta u + M_\eta v)\, n\log(2M_\eta n) \\
&\quad + M_\eta u n\log(4n) + 6nM_\eta u + 2n^{1/2}\left|\alpha_n^-(1,v)\right| + 2n^{1/2}\left|\alpha_n^-(1,1)\right| \, d\Gamma_{Sn}(u)\,dv \\
&\le \left(3M_\eta n^{5/8}\log(4M_\eta n) + 6M_\eta n^{5/8} + 2n^{1/2}\sup_{0\le v\le 1}\left|\alpha_n^-(1,v)\right| + 2n^{1/2}\left|\alpha_n^-(1,1)\right|\right) \\
&\quad \cdot \int_0^{n^{-3/8}}\int_0^{n^{-3/8}} d\Gamma_{Sn}(u)\,dv \;\to\; 0.
\end{aligned}$$

The region $C_{\varepsilon,1,2}$ can be treated in a similar way as $C_{\varepsilon,1,1}$.

For $C_{\varepsilon,3,1}$ and $C_{\varepsilon,3,2}$ we use $|\log(1+x)-x| \le x^2$, for $x \ge -0.5$, and the convergence in probability of P_n/P uniform over certain rectangles (the $A_j^{\pm}$) to 1. This follows from, e.g., Einmahl (1987), Inequality 2.9 or Theorem 3.3. Then, with probability tending to 1,

$$\begin{aligned}
|\log R(Q_S(u),Q(v))| \;\le\;& \frac{\left[v\alpha_n^-(u,1) - uv\alpha_n^-(1,1) - \alpha_n^-(u,v) + u\alpha_n^-(1,v)\right]^2}{\frac{1}{2}u(1-u)v(1-v)} \\
&+ \frac{\left[v\alpha_n^+(u,1) - uv\alpha_n^+(1,1) - \alpha_n^+(u,v) + u\alpha_n^+(1,v)\right]^2}{\frac{1}{2}u(1-u)v(1-v)} \\
&+ \frac{\left[\alpha_n^-(1,v) - v\alpha_n^-(1,1)\right]^2}{\frac{1}{2}v(1-v)} + \frac{\left[\alpha_n^+(1,v) - v\alpha_n^+(1,1)\right]^2}{\frac{1}{2}v(1-v)} \\
&+ \frac{\left[\alpha_n^+(u,1) - u\alpha_n^+(1,1) - \alpha_n^-(u,1) + u\alpha_n^-(1,1)\right]^2}{u(1-u)} + 4\alpha_n^-(1,1)^2 \\
\le\; & 4\frac{\alpha_n^-(u,v)^2 + v^2\alpha_n^-(u,1)^2 + u^2\alpha_n^-(1,v)^2 + u^2v^2\alpha_n^-(1,1)^2}{\frac{1}{2}u(1-u)v(1-v)} \\
&+ 4\frac{\alpha_n^+(u,v)^2 + v^2\alpha_n^+(u,1)^2 + u^2\alpha_n^+(1,v)^2 + u^2v^2\alpha_n^+(1,1)^2}{\frac{1}{2}u(1-u)v(1-v)} \\
&+ 2\frac{\alpha_n^-(1,v)^2 + v^2\alpha_n^-(1,1)^2}{\frac{1}{2}v(1-v)} + 2\frac{\alpha_n^+(1,v)^2 + v^2\alpha_n^+(1,1)^2}{\frac{1}{2}v(1-v)} + 2\alpha_n^+(1,1)^2 \\
&+ 4\frac{\alpha_n^+(u,1)^2 + u^2\alpha_n^+(1,1)^2 + \alpha_n^-(u,1)^2 + u^2\alpha_n^-(1,1)^2}{u(1-u)} + 2\alpha_n^-(1,1)^2 \\
\le\; & 32\left[\frac{\alpha_n^-(u,v)^2 + \alpha_n^+(u,v)^2}{uv} + \frac{\alpha_n^-(u,1)^2 + \alpha_n^+(u,1)^2}{u} + \frac{\alpha_n^-(1,v)^2 + \alpha_n^+(1,v)^2}{v}\right.
\end{aligned}$$

$$+ 2\alpha_n^-(1,1)^2\Big].$$

Theorem 3.1 in Einmahl (1987) yields, for either choice of sign,

$$\sup_{0<u,v\le 1} \frac{|\alpha_n^{\pm}(u,v)|}{(uv)^{1/4}} = O_P(1).$$

Hence we find

$$\begin{aligned}
&\iint\limits_{C_{\varepsilon,3,1}\cup C_{\varepsilon,3,2}} |\log R(Q_S(u), Q(v))|\; d\Gamma_{Sn}(u)\, dv \\
&= O_P(1) \cdot \iint\limits_{C_{\varepsilon,3,1}\cup C_{\varepsilon,3,2}} \left(\tfrac{1}{\sqrt{uv}} + \tfrac{1}{\sqrt{u}} + \tfrac{1}{\sqrt{v}} + 1\right) d\Gamma_{Sn}(u)\, dv \\
&= O_P(1) \cdot \iint\limits_{C_{\varepsilon,3,1}\cup C_{\varepsilon,3,2}} \tfrac{1}{\sqrt{uv}} d\Gamma_{Sn}(u)\, dv \;=\; O_P(\sqrt{\varepsilon}),
\end{aligned}$$

uniformly in ε, because of (5.7). This completes the proof of (5.4). □

Bibliography

Anderson, T.W., and Darling, D.A. (1952), "Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes," *Annals of Mathematical Statistics*, 23, 193–212.

Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H., and Tukey, J.W. (1972), *Robust Estimation of Location: Survey and Advances*, Princeton University Press, Princeton.

Antonidiadis, A., and Gijbels, I. (2002), "Detecting Abrupt Changes by Wavelet Methods," *Journal of Nonparametric Statistics*, 14, 7–29.

Baringhaus, L. (1991), "Testing for Spherical Symmetry of a Multivariate Distribution," *The Annals of Statistics*, 19, 899–917.

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004), *Statistics of Extremes: Theory and Applications*, Wiley, New York.

Billingsley, P. (1968), *Convergence of Probability Measures*, Wiley, New York.

Carlstein, E. (1988), "Nonparametric Change-Point Estimation," *The Annals of Statistics*, 16, 188–197.

Chaudhuri, P., and Marron, J.S. (1999), "SiZer for Exploration of Structures in Curves," *Journal of the American Statistical Association*, 94, 807–823.

Chmielewski, M.A. (1981), "Elliptically Symmetric Distributions: A Review and Bibliography," *International Statistical Review*, 49, 67–74.

Cleveland, W.S. (1979), "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association,* 74, 829-836.

Cobb, G.W. (1978), "The Problem of the Nile: Conditional Solution to a Change-Point Problem," *Biometrika*, 64, 243–251.

Crawford, M.L.L., Taylor, M.S., and Tingey, H.B. (1990), "Quantiles of the Anderson-Darling Statistic," *Communications in Statistics - Simulation and Computation*, 19, 1007–1014.

Csörgő, M. and Révész, P. (1978), "Strong Approximations of the Quantile Process," *The Annals of Statistics*, 6, 882–894.

Dempfle, A., and Stute, W. (2002), "Nonparametric Estimation of a Discontinuity in Regression," *Statistica Neerlandica*, 56, 233–242.

Dümbgen, L., and Walther, G. (2008), "Multiscale Inference about a Density," *The Annals of Statistics*, 36, 1758–1785.

Einmahl, J.H.J. (1987), *Multivariate Empirical Processes*, CWI Tract 32. Centre for Mathematics and Computer Sciences, Amsterdam.

Einmahl, J.H.J., Gantner, M. and Sawitzki, G. (2010), "The Shorth Plot," *Journal of Computational and Graphical Statistics*, 19, 62–73.

Einmahl, J.H.J., Gantner, M. and Sawitzki, G. (2010), "Asymptotics of the Shorth Plot," *Journal of Statistical Planning and Inference*, 140, 3003–3012.

Einmahl, J.H.J., and Mason, D.M. (1992), "Generalized Quantile Processes," *The Annals of Statistics*, 20, 1062–1078.

Einmahl, J.H.J., and McKeague, I.W. (2003), "Empirical Likelihood Based Hypothesis Testing," *Bernoulli*, 9, 267–290.

Fang, K.-T., Kotz, S., and Ng, K.W. (1990), *Symmetric Multivariate and Related Distributions*, Chapman and Hall, London.

Fang, K.-T., and Liang, J.J. (1999), "Testing Spherical and Elliptical Symmetry". In S. Kotz, C.B. Read, D.L. Banks (Eds.), *Encyclopedia of Statistical Sciences* (Update), 686–691 (Vol 3), Wiley, New York.

Gijbels, I., and Goderniaux, A.-C. (2004), "Bandwidth Selection for Change-point Estimation in Nonparametric Regression," *Technometrics*, 46, 76–86.

Gijbels, I., Hall, P. and Kneip, A. (1999), "On the Estimation of Jump Points in Smooth Curves," *Annals of the Institute of Statistical Mathematics*, 51, 231–251.

Gijbels, I., Lambert, A., and Qiu, P. (2007), "Jump-Preserving Regression and Smoothing Using Local Linear Fitting: a Compromise," *Annals of the Institute of Statistical Mathematics*, 59, 235–272.

Govindarajulu, Z. (2007), *Nonparametric Inference*, World Scientific Publishing, Toh Tuck Link (Singapore).

Grégoire, G., and Hamrouni, Z. (2002), "Change Point Estimation by Local Linear Smoothing," *Journal of Multivariate Analysis*, 83, 56–83.

Grübel, R. (1988), "The Length of the Shorth," *The Annals of Statistics*, 16, 619–628.

Haldane, J.B.S. (1948), "A Note on the Median of a Multivariate Distribution," *Biometrika*, 35, 414–5.

Hall, P., and Titterington, D.M. (1992), "Edge-Preserving and Peak-Preserving Smoothing," *Technometrics*, 34, 429–440.

Horváth, L., and Kokoszka, P. (2002), "Change-Point detection With Non-parametric Regression," *Statistics*, 36, 9–31.

Horváth, L., Kokoszka, P., and Steinebach, J. (1999), "Testing for Changes in Multivariate Dependent Observations With an Application to Temperature Changes," *Journal of Multivariate Analysis*, 68, 96–119.

Hyndman, R.J. (1996), "Computing and Graphing Highest Density Regions," *The American Statistician*, 50, 120–126.

Hyndman, R.J., Bashtannyk, D.M., and Grunwald, G.K. (1996), "Estimating and Visualizing Conditional Densities," *Journal of Computational and Graphical Statistics* 5, 315–336.

Kim, C.S., and Marron, J.S. (2006), "SiZer for Jump Detection," *Journal of Nonparametric Statistics*, 18, 13–20.

Koltchinskii, V.I, and Li, L. (1998), "Testing for Spherical Symmetry of a Multivariate Distribution," *Journal of Multivariate Analysis*, 65, 228–244.

Kraus, E.B. (1955), "Secular Changes of Tropical Rainfall Regimes," *Quarterly Journal of the Royal Meteorological Society*, 81, 198–210.

Kvam,P.H., and Vidakovic B. (2007), *Nonparametric Statistics with Applications to Science and Engineering*, Wiley, New York.

Laurent , A.G. (1974), "The Intersection of Random Spheres and the Noncentral Radial Error Distribution for Spherical Models," *The Annals of Statistics*, 2, 182–189.

Liang, J., Fang, K.-T., Hickernell, F.J. (2008), "Some Necessary Uniform Tests for Spherical Symmetry," *Annals of the Institute of Statistical Mathematics,* 60, 679–696.

Loader, C.R. (1996), "Change Point Estimation Using Nonparametric Regression," *The Annals of Statistics*, 24, 1667–1678.

Loader, C. (1999), *Local Regression and Likelihood*, Springer, New York.

McDonald, J.A., and Owen, A.B. (1986), "Smoothing with Split Linear Fits," *Technometrics*, 28, 195–208.

McNeil, A.J., and Nešlehovà, J. (2009), "Multivariate Archimedean Copulas, d-monotone Functions and l_1-norm Symmetric Distributions", *The Annals of Statistics*, 37, 3059–3097.

Minnotte, M.C., and Scott, D.W. (1993), "The Mode Tree: A Tool for Visualization of Nonparametric Density Features," *Journal of Computational and Graphical Statistics*, 2, 51–68.

Müller, D.W., and Sawitzki, G. (1991), "Excess Mass Estimates and Tests for Multimodality," *Journal of the American Statistical Association*, 86, 738–746.

Müller, H.-G. (1992), "Change-Points in Nonparametric Regression Analysis," *The Annals of Statistics*, 20, 737–761.

Owen, A.B. (2001), *Empirical Likelihood*, Chapman and Hall/CRC, Boca Raton.

Park, C., and Kim, W.-C. (2006), "Wavelet Estimation of a Regression Function With a Sharp Change Point in a Random Design," *Journal of Statistical Planning and Inference*, 136, 2381–2394.

Racine, J.S. (2009), "Nonparametric and Semiparametric Methods in R," *Nonparameetric Econometric Methods*, 25, 335–375.

Sánchez-Borrego, I.R., Martínez-Miranda, M.D., and González-Carmona, A. (2006), "Local Linear Kernel Estimation of the Discontinuous Regression Function," *Computational Statistics*, 21, 557–569.

Sawitzki, G. (1994), "Diagnostic Plots for One-Dimensional Data," In: R. Ostermann and P. Dirschedl, ed., *Computational Statistics, 25th Conference on Statistical Computing at Schloss Reisensburg*, Physica-Verlag/Springer, Heidelberg, 237–258.

Sawitzki, G. (2009), *Computational Statistics: An Introduction to R*, Chapman & Hall/CRC Press, Boca Raton (FL).

Schmitz, H. P., and Marron, J.S. (1992), "Simultaneous Estimation of Several Size Distributions of Income," *Econometric Theory*, 8, 476–488.

Serfling, R.J. (2006) "Multivariate Symmetry and Asymmetry". In S. Kotz, N. Balakrishnan, C.B. Read, B. Vidakovic (Eds.), *Encyclopedia of Statistical Sciences*, 5338–5345 (2nd Ed.), Wiley, New York.

Shorack, G.R., and Wellner, J.A. (1986), *Empirical Processes with Applications to Statistics*, Wiley, New York.

Silverman, B.W. (1981), "Using Kernel Density Estimates to Investigate Multimodality," *Journal of the Royal Statistical Society*, 43, 97–99.

Silverman, B.W. (1986), *Density Estimation*, Chapman and Hall, London.

Smith, P.J. (1977), "A Nonparametric Test for Bivariate Circular Symmetry Based on the Empirical Cdf," *Communications in Statistics - Theory and Methods*, 6, 209–220.

Szabłowski, P.J. (1998), "Uniform Distributions on Spheres in Finite Dimensional L_α and Their Generalizations," *Journal of Multivariate Analysis*, 64, 103–117.

Theus, M., and Urbanek, S. (2009), *Interactive Graphics for Data Analysis: Principles and Examples*, Chapman & Hall/CRC Press, Boca Raton (FL).

du Toit, S.H.C., Steyn, A.G.W., and Stumpf, R.H. (1986), *Graphical Exploratory Data Analysis,* Springer, New York.

Tufte, E.R. (1983), *The Visual Display of Quantitative Information,* Graphics Press, Cheshire (CT).

Tukey, P.A., and Tukey, J.W. (1981), "Data-Driven View Selection; Agglomeration and Sharpening," in *Interpreting Multivariate Data* ed. V. Barnett, Wiley, New York.

van der Vaart, A.W. (1998), *Asymptotic Statistics,* Cambridge University Press, New York.

Vervaat, W. (1972), "Functional Central Limit Theorems for Processes with Positive Drift and their Inverses," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 23, 245–253.

Wang, Y. (1995), "Jump and Sharp Cusp Detection by Wavelets," *Biometrika*, 82, 385–397.

Wolfowitz, J. (1942), "Additive Partition Functions and a Class of Statistical Hypotheses", *The Annals of Mathematical Statistics*, 13, 247–279.

Wu, J.S., and Chu, C.K. (1993), "Kernel-Type Estimation of Jump Points and Values of a Regression Function," *The Annals of Statistics*, 21, 1545–1566.